Activity Report Rennes - Bretagne Atlantique 2016

Edition: 2017-06-08

List of Inria's Research Teams

Team ANJA

ANJA

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

RESEARCH CENTER
Rennes - Bretagne-Atlantique

THEME
Optimization, machine learning and statistical methods

Table of contents

Team ANJA

Creation of the Team: 2016 January 01, end of the Team: 2016 December 31

Keywords:

Computer Science and Digital Science:
3.4.1. - Supervised learning
3.4.2. - Unsupervised learning
3.4.4. - Optimization and learning
3.4.5. - Bayesian methods
6.1.2. - Stochastic Modeling (SPDE, SDE)
6.1.3. - Discrete Modeling (multi-agent, people centered)
8.2. - Machine learning

Other Research Topics and Application Domains:
9.5.2. - Juridical science
9.5.3. - Economy, Finance
9.5.5. - Sociology
9.5.6. - Archeology, History

1. Members

Research Scientist
Jacques Levy Vehel [Team leader, Inria, Researcher, HDR]

Faculty Member
Anne Philippe [Univ. Nantes, Professor, HDR]

Technical Staff
Jerome Dupre [Inria]
Marie-Anne Vibet [Univ. Nantes]

Post-Doctoral Fellow
Yue Huang [Inria]

Administrative Assistant
Anne Claire Binetruy [Inria]

Other
Claire Crouzilles [Inria, Intern, from Jun 2016 until Jul 2016]

2. Overall Objectives

2.1. Overall Objectives

Applications of mathematics in social sciences (economy, sociology, law,...) are increasing at a fast pace. In a growing number of areas, decisions are being made in an *actuarial* way rather than in a *clinical* one, as has been the norm for decades or even centuries. This means that actions are based on the outputs of (mathematical) models instead on the informed judgement of experts. This occurs daily in financial economics (for instance for option pricing, portfolio management and risk assessment), and is also on the verge of becoming routine in certain domains of law and certain countries (most notably the USA, where examples include evidence-based sentencing and the setting of certain fines).

Such a paradigm shift certainly has several benefits, but it may also be harmful at least because it *rigidifies* the system. What we mean by this is that, when most actors in a given field use the same model, with possibly minor variants, massive unanticipated effects will occur:

1. if the model is either wrong, or wrongly specified or used beyond its nominal conditions, the fact that everyone makes the same erroneous decisions may lead either to system-wide crises (this has happened several times in finance in recent years) or to profound uncontrolled modifications of reality (this has also occurred in finance but also in law). This second case, leading to “silent revolutions”, is maybe more problematic;
2. even when the model is approximately correct or used in normal conditions, hard and fast, uniform answers will reduce diversity in an often damaging way.

Both situations above are related to the concept of *performativity* introduced by social scientists. More precisely, item 1 leads to what we call convergent or divergent performativity, while item 2 leads systematically to convergent performativity.

How can one eliminate or at least reduce the negative impacts of the generalizing use of actuarial approaches in social sciences ? Clearly, the trend seems unstoppable and even if it were, this would probably not be desirable: quantitative, mathematically well-founded methods are likely to permit great progress in many domains. There is for instance a strong need for good models that would help judges fix financial amounts in various areas including spousal support and fines in cases of violation of intellectual property. The aim of Anja is to propose ways to minimize the drawbacks of model-based approaches using essentially the following devices:

1. be aware of the performative effect of models: by taking into account the impact of actuarially-based actions on reality when setting-up a model, this impact will be endogenised and, thus, to some extent, controlled. In other words, whenever possible, we study the feedback between reality and the proposed model.
2. insist that the output of models be systematically probabilistic: instead of a hard answer, always propose a range of answers with associated probability distribution. This should greatly reduce the rigidification of the system by allowing to re-introduce a part of expertise, thus preserving some diversity, and in turn minimise the occurrence of crises or unwanted modifications of reality. From a theoretical point of view, this amounts to using tools introduced in the now well-developed area of *uncertainty propagation*. The basic idea is that uncertainties on the various parameters are propagated at all stages of the modelling process, so that, at each step, computations are made on probability densities rather than numbers. This approach raises various difficulties when the models are non-linear, as will typically be the case in our applications.

Although our long term aim is to provide a general frame enabling to fulfil the above objectives, our first studies will focus on a limited number of well identified situations in economy and law described in Sections 4.1 and 4.2.

Our program clearly requires strong interdisciplinary collaborations. Anja is actively involved in on-going work with financial economists and lawyers.

3. Research Program

3.1. Research Program

The aim of Anja is to develop mathematical models in selected areas of SHS, which include, at this time, economy/finance, law, and archaeology. These models are essentially probabilistic. This entails that our theoretical studies mainly lie in the fields of probability and statistics.

A major focus of Anja is on providing mathematical analyses of how performativity operates in economy/finance and law, where performativity is understood as the phenomenon by which applying models co-constructs a new reality by the very fact that the existing reality was not properly apprehended. We are chiefly concerned with performativity that results from mathematical models. Indeed, while performativity exists before and independently of such models, mathematics may, and already have, strengthened the performative power in a significant way. This has occurred so far in a uncontrolled fashion, and thus in a typically damaging way. The essence of our work is to shed light on the mechanisms mediating the increase of performativity brought by mathematical models, thus allowing one to manage their effects and hopefully orient them towards an improvement of the reality they transform.

We stress the important fact that this program allows us to go well beyond what is typically achieved in sociological studies. For instance, many such studies have evidenced the role of performativity in the context of financial theory and how it shapes today's markets [46]. This is certainly useful in order to exert political control on the tools proposed by, e.g., economists. However, such an exogenous control is not fully satisfactory because it does not provide explicit procedures to enhance the models. This is due to the fact that these studies have not permeated the technical literature to a point where it would have a significant impact on the definition and practical use of models. One explanation for this is that, though convincing, these analyses do not provide mathematical or otherwise applicable tools to modify practices: they explain general mechanisms through which, for instance, economics performs the reality of economies, but do not shed light on the precise mathematics that mediate these mechanisms. We believe that it is important that mathematicians tackle this issue. In other words, we think that it is extremely useful to reverse that statement made in [48]: *"en souscrivant au programme de la performativité, la recherche en sociologie économique ne se contente plus de partager ses objets d'études avec les sciences économiques, elle inclut ces dernières dans ses propres objets d'étude"* [0]. Our "reverse" statement is that economical sciences, and, more generally, mathematics applied in SHS, should include in their research objectives the sociological impact they create through performativity: mathematicians need to take into account in their models the fact that reality will be transformed by them, and thus model also this transformation.

The first goal of Anja is precisely to fill this gap, that is, to pinpoint which parts of a given mathematical model are responsible for performative effects and how this occurs. It is important to stress that, in our view, this means that the performativity of mathematical models will be itself assessed with mathematical tools. This endogeneisation permits to measure quantitatively the impact of models on reality. In turn, this quantification opens the way to our second and more ambitious goal: indeed, we propose paradigms allowing one to control the performativity of mathematical models. The main mechanisms we use in that view are as follows:

- systematically take into account the fact that the models will perform reality. This means that, when defining our models, we try as much as possible to foresee how applying them will transform practices, and then adjust them in such a way that these modifications are desirable and under control;
- impose that the output of our models always be *probability distributions* rather than hard prescriptions. In other words, recognizing that modelling in SHS, in addition to being typically extremely complex with a large number of variables and with many sources of errors in the process of calibrating the parameters, always involves addressing a moving reality that will be transformed by the very application of the model, we insist that, at all stages of the analysis, *uncertainties* be propagated, as is routinely done in industrial fields such as aeronautics, so that the answer of the system will be probabilistic. We use in particular Bayesian analysis in that view, in order to incorporate information on unknown parameters using prior probability distributions that are sequentially updated after the acquisition of each new observation.

As a longer term perspective, we intend to propose a general mathematical model of performativity. The current literature has already proposed general analyses of mechanisms through which theories can become performative: most notably, [35] has identified three main such channels, namely institutional design, social

[0] "By subscribing to the performativity program, research in economical sociology goes beyond sharing its objects of study with economical sciences: it includes the latter in its own objects of study.

norms, and language, as well as the way in which culture and accountability affect their course of operation. Our aim will be more focused: we will concentrate on the sole performativity of mathematical models, but in this restricted frame, we wish to propose quantitative, mathematical analyses. In other words, a mathematical model of performativity should allow one to answer questions such as: when can one expect that a theory is likely to be performative, what exactly are the conditions favouring performativity, which indices should one look for in order to detect a performative influence, how can one predict whether performativity will be convergent or divergent, which aspects of reality a theory will affect and how, and finally what are the means to minimize its negative effects.

One important motivation of Anja is that we feel that, as researchers in mathematics, we are partly responsible for the way mathematics is used in social sciences. In particular, while we strongly believe that mathematical models have already and will continue to enhance our social lives as they have improved our understanding and control in natural sciences, extra caution is needed because of their performative power explained above. The core of our work is that such caution can be exercised (a) by recognizing that models impact reality and by taking into account their performative power in their very definition, and (b) by using systematically probability and statistics: in a nutshell, imposing that mathematical answers to questions in social sciences always take the form of a probability distribution should (1) remind users that no mathematical model is able to provide a definitive hard and fast answer when it comes, e.g., to computing the amount of a fine in competition law, and (2) allow one to tame the inherent complexity of human-related matters, thus providing useful guides for making informed decisions.

Of course, we do not address performativity issues in all social and human sciences. Rather, we focus on two domains where we already have an expertise: economy/finance, and law. Details on our studies in these fields are given in sections 4.1 and 4.2.

Our program cannot be realistically realised without strong collaborations from specialists in the SHS fields we deal with. In law, our expertise is brought by Jérôme Dupré, which holds a Ph.D. in law and is also a former judge. As far as archaeologyis concerned, we collaborate with Philippe Lanos (senior researcher at CNRS). He is an expert in archaeomagnetism and its applications to materials dating in archaeology.

4. Application Domains

4.1. Economy and finance

4.1.1. Basel III and Solvency 2 regulations

As amply demonstrated above, economy is a field where the performativity of mathematical models is particularly noticeable. This has become even more so in recent years in finance because international regulations have fundamentally changed since the Basel II Accords. Among other evolutions, Basel II and III explicitly impose that computations of capital requirements be model-based. The same is true of the Solvency 2 directive, a European regulation aiming in particular at evaluating the amount of capital that insurance companies must hold to reduce the risk of insolvency, much in the spirit in the Basel Accords.

This paradigm shift in risk management has been the source of strong debates among both practitioners and academics, who question whether such model-based regulations are indeed more efficient.

A common feeling in the industry is that regulations will sometimes give a false impression of security: risk managers tend to think that a financial company that would fulfil all the criteria of, say, the Basel III Accords on capital adequacy, is not necessarily on the safe side. This is so mainly because many risks, and most significantly systemic or system-wide risks, are not properly modelled, and also because it is easy to manipulate to some extent various risk measures, such as Value at Risk (VaR).

In parallel, a fast growing body of academic research provides various arguments explaining why current regulations are not well fitted to address risk management in an adequate way, and may even, in certain cases, worsen the situation. In other words, they have a divergent performativity effect.

Our first angle to tackle the performativity of these regulations is to question the Gaussian assumption that is implicitly made in designing them. More precisely, we have already shown in [11], [12] that, in some situations, and because of this assumption, prudential rules are themselves the source of a systemic risk. In [12], it was explained how a wrong model of price dynamics coupled to the regulatory VaR constraint tends to systematically increase Tail Conditional Expectation. [11] details how trying to minimize VaR under Gaussian beliefs for the dynamics of returns when actual movements are stable non-Gaussian results in fact in maximization of VaR. Along with the concept of endogenous risk put forward in [44], this body of work provides a mathematical description of how models perform financial reality: this is a perfect example of divergent performativity, since, because of a wrong model, (mandatory) actions are taken that make financial markets even less similar to the model. More technically, assume the simplest model of returns movements, that is, Brownian motion. Brownian motion is the symmetric stable motion characterized by the stability index $\alpha = 2$ and a given scale parameter σ^0. Under reasonable assumptions, minimizing VaR in a Brownian market amounts to minimizing the variance. However, in a stable market where $\alpha < 2$, which therefore is subject to jumps, minimizing VaR requires to maximize α while choosing an intermediate value of σ. Furthermore, actions taken under a Brownian belief will tend not only to minimize σ but also α: therefore, implementing VaR-based regulations founded on the wrong Brownian model tends to decrease α, making the market even "more" non-Brownian. This is exactly the definition of divergent performativity.

The work in [11], [12] is only one possible mechanism of performativity, although maybe the simplest one. Starting from this, one may progress in two directions: propose regulations that will avoid at least the particular kind of performativity just described, and study more complex models and their performative effects.

As for the first direction, assuming a stable non-Brownian market, we need to understand what kind of constraints would lead to actions favouring an increase rather than a decrease of α. Our first idea is to explore counter-cyclical measures, as current regulations are often blamed for their pro-cyclical effect. In a nutshell, pro-cyclicity is entailed by the fact that, in market downs, actors will be forced by regulations to reduce their exposure, thus amplifying downwards movements. We plan to investigate how this translates into modifications of the (α, σ) couple, and check whether basing regulations on the time evolution of this couple would be efficient. For instance, one might imagine measuring (α, σ) as a function of time, and let financial companies increase or decrease their solvency capital requirements based on the coupled evolution.

As for the second direction, we remark that, since regulations tend to endogenously modify both volatility and jump intensity, it seems natural to define and study processes where the local regularity varies in time, possibly in relation with the value of the process. We have introduced such classes of processes in recent years. We plan to deepen their study in the light of their possible adequacy for the mathematical modelling of performativity. We briefly describe now the first actions we will take in this respect.

4.1.2. Multistable and self-stabilizing processes for financial modelling

It is widely accepted that the dynamics of most financial instruments display jumps and there is a huge literature dealing with jump processes in all areas of financial engineering [32]. In order to get a better understanding of these dynamics, we have developed in recent years various instances of *multistable processes*. These processes were introduced in [4] and further studied e.g. in [8]. Their main feature is that their local intensity of jumps varies in time. In view of their application, we plan to study the following points:

- Recognizing that the local characteristics (intensity of jumps and scale) vary in time implies that evolution equations these parameters must be proposed for these parameters. We have started to develop Hull and White-like models, where auxiliary EDS are satisfied by both scale and the intensity of jumps. This will hopefully allow one to model in a satisfactorily manner implicit volatility surfaces.

[0] recall that a stable motion is a process with independent and identically distributed increments, where each increment follows a stable law $S_\alpha(\sigma, \beta, \mu)$. The parameter $\alpha \in (0, 2]$ characterizes the jump intensity - the smallest α, the largest the jump intensity, with no jumps when $\alpha = 2$, that is, for Brownian motion -, σ is the scale parameter - proportional to the variance when $\alpha = 2$ -, β is the skewness parameter and μ the location one.

- Robust statistical estimation of $\alpha(t)$ (or of the couple $(\alpha(t), h(t))$ in the case of the so-called linear multifractional multistable motion) is necessary. Some results are presented in [45], but other methods should be studied.
- Self-regulating processes are processes where the local regularity is a function of the amplitude. They were introduced in [1] and further studied e.g. in [3]. It seems natural to follow the same approach and define "self-stabilizing processes" as processes where the local index of stability is a function of the amplitude. Certain tools used for defining some SRP, namely the fixed point theorem, could be adapted, with the difference that the underlying space will not be the one of continuous functions, but the one of càdlàg functions. As a consequence, the Prohorov metric may have to be considered instead of the sup-norm. We have some preliminary results in this direction, which also include the definition of Markovian self-stabilizing processes. Statistical issues (that is, the estimation of the "self-stabilizing" function) need also be addressed.

4.1.3. Multifractional and self-regulating processes for financial modelling

Besides multistable motions, we will also continue to investigate the use of multifractional Brownian motion in financial modelling. Previous works [29] have shown the potential of this approach, in particular for reproducing certain features of the volatility process [51], and we plan to pursue this line of study. More precisely, we will investigate the following matters:

- The instance of self-regulating processes built so far [1] are not progressive, in the sense that paths are constructed globally rather than in a chronological manner. For this reason, they do not provide adequate models for time series encountered in economy and finance. We will put some effort in trying to construct progressive self-regulating processes. Our first attempts will be based on pathwise stochastic integrals as well as on Skohorod integrals.
- Once progressive self-regulating processes have been built and their basic probabilistic properties been investigated, the second step will consist in constructing estimators for the self-regulating function (that is, the function relating amplitude and regularity). This is of course essential for applications.
- We will finally investigate precisely which economical or financial times series display self-regulation, and examine the performative effect of current regulations when such models are in force.

4.1.4. Performativity of monetary policies

It seems clear that, besides prudential regulations, monetary policies such as quantitative easing used by central banks in Europe, Japan and the USA have a strong impact on economy [0]. There is already a huge literature studying this impact. From a broader perspective, many actions taken by financial authorities are designed in a conceptual frame where volatility is all there is to risk. We believe that incorporating at least another dimension related to jumps is essential for proper control. In this respect, we plan to analyse in a quantitative way what is the impact on the stability of markets of the various measures taken by central banks in recent years, such as Zero Interest Rates Policies, Large Scale Assets Purchases, Forward Guidance or Long Term Refinancing Operations, when one takes into account the jump dimension of risk. Such measures have led to typically very low volatility on the markets. But, as C. Borio of BIS recently stated [30], "history teaches us that low volatility and risk premia are not the signs of smaller risk, but rather than investors are ready to take large risks. The less investors fear risk, the more dangerous the situation is". In other words, recent monetary policies seem to have lowered volatility at the expense of increasing the intensity of jumps. This view is supported by a number of studies in recent years by the BIS. For instance, [26] argues that the accommodative monetary policy have pushed volatility to low levels in various ways: directly by reducing the amplitude of interest rate movements and by removing to a large extent uncertainty about interest rate changes; and indirectly because an environment of low yields on high- quality benchmark bonds favours risk-taking. Investors then tend to have a lower perception of risk, and thus be inclined to take riskier positions.

[0] In a nutshell, quantitative easing is an unconventional monetary policy by which central banks create new money to buy financial assets in view of stimulating the economy.

Studying such a performative effect is typically in the focus ofAnja. Our first attempts in this direction will be again to use stable or multistable processes in place of the Brownian motion as a source of randomness. The obvious approach is to rewrite current models with this modification. This will however require to define several new notions adapted to this situation. More precisely, most computations in classical models crucially depend on the fact that all the quantities involved are square integrable, a property not available when one deals with (multi-)stable processes. As a consequence, correlations, for instance, are not well-defined; this is a problem as they serve as a fundamental tool in such studies. One possible way out would be to use CGMY or other tempered stable processes instead of stable ones, since this would bring us back in the realm of L^2 random variables. The price to pay is that we lose stability, meaning that aggregate behaviours are more difficult to assess. A more ambitious but potentially more fruitful approach is to to start again from the modified classical models but to extend their study in a stable frame so as to be able to compute joint distributions.

Another, very different path, is to use the mathematical theory of causality to tackle these questions [49]. We will recall in the next section some facts about causality. Recent studies have tried to tackle the question of determining the causal structure among economic quantities. For instance, results in [33] suggest that per capita real balances and real per capita private gross domestic product are both causes of real per capita consumption expenditures and that real per capita consumption expenditures and real per capita private gross domestic product in turn cause real per capita gross private domestic fixed investment in a four-variables vector autoregressive model of US macro-economic data for the period January 1949 to April 2002. We plan to use both constraint-based methods and Bayesian approaches to study the causal structure in a graph where the nodes are the various quantities manipulated by quantitative easing policies. As always, one of the main problems will be to define the set of sufficient variables.

4.2. Law

There are now many ways in which mathematics are applied to law. They include the following approaches:

1. the classical domain of *Law and Economics*
2. the more recent statistical approaches
3. approaches using tools of mathematical logic.

Given our expertise, we are concerned with approaches 1 and 2: our first applications are based either on a mix of economic and statistical methods, or on purely statistical ones. We will also develop original probabilistic models.

From a general point of view, the benefits of using actuarial models in law is twofold:

- mathematical models should allow for a more profound understanding of law structures and rules. Indeed, as explained in [47], law can be seen as an information technology in the sense that it provides information to the community about the content of legal norms and, in its common law form, elicits information about the world from the disputes before a court. In this two-way path, tension between law's potential for certainty and its capacity for discovery reflects in part the imperfect circulation of information. The joint use of adequate mathematical models and big data tools should greatly enhance this circulation, thus improving the efficiency of the system as a whole;
- in a more complex and more informed world, legal procedures are likely to become more frequent. However, the state resources devoted to law cannot increase without bounds. Making available tools that would facilitate amicable settlement is then of strong interest. In particular, models allowing one to estimate outputs of legal decisions, at least in certain areas and in a rough way, would certainly draw people to be more inclined to negotiate rather than go to court, thus reducing the burden put on the legal system. This tendency is already quite noticeable in particular in the USA, where so-called *on-line dispute resolution systems* gain popularity.

We contribute to both these goals, paying in addition extra caution to the performative aspects. Our first studies are detailed in the next sections.

4.2.1. Law-Mathematics correspondences

In order to root our subsequent studies on firm bases, we intend to start by evidencing some parallel notions in law and mathematics, and to study if they are profound enough to yield useful tools. While this will inevitably be sometimes rather qualitative, it will definitely shed some light on how to model legal reasoning in a mathematical way.

An example of such a qualitative link is the fact judges, as mathematicians, when faced with a question, often have immediately a intuition of their answer. In a second phase, lawyers try to find which legal texts or jurisprudence allow them to justify this answer, while mathematicians invoke a series of computations and known theorems to do the same. In both cases, if no path is found to the initial answer (that is, no legal texts or no valid sequence of computations), the practitioner tries to defend or prove the opposite one. We have no idea yet how to formalize this parallelism, but this will be a topic of study. More quantitative ones are the following:

1. Weights and linear models

   Judges often say that they weigh different factors when they need to make a decision. The obvious corresponding mathematical notion is the one of linear models, where variables are linearly combined to produce an output. We will choose some simple domains, such as for instance child support, to check whether the decided amount is indeed obtained by weighting the criteria that judges are supposed to take into account.

   This requires to analyse a large amount of case law and assessing the fit of various linear or generalised linear models. State-of-the-art techniques in machine learning are used in this connection.

2. Causality

   Finally, an obvious and probably fruitful correspondence between both domains rests on the notion of causality. Determining which events are causes of others is clearly a crucial task in courts, since evidencing responsibilities is at the core of making informed judgements.

   On the other hand, statisticians have, until rather recently, avoided to consider causal questions, concentrating on correlations. This is still true today, where most researchers and practitioners would claim that statistics can only evidence dependencies between random variables but cannot assess causal links, except when controlled experiments may be performed. It is hard to think of a situation in law where one could perform such experiments.

   However, a growing community has started to develop what now seems to be a somewhat coherent theory, termed causality theory, that allows one to efficiently decide if a variable X is indeed a cause of a variable Y under some conditions [49]. Apart from theoretical developments, this theory has been applied in various domains, and most notably in economy and biomedical studies. We are not aware of any applications in law.

   We study this area in two ways:

   – the most direct one is to choose a specific domain, analyse some decisions in it in light of the legal and jurisprudential criteria that are supposed to base them, and check whether they are indeed causes of the decision in the sense of causality theory. More generally, we try to construct the whole Bayesian network associated with a given field;

   – a more ambitious goal is to question whether the way law sees and organizes causality is anything like what is performed in statistical causality theory. This task requires an abstract model of legal causality that must be constructed from scratch. This is a long term aim.

4.2.2. Scales and performativity

We have just won a call “Droit, justice et numérique” of the “Mission de recherche Droit et Justice”, a “groupement d’intérêt public” created by the French ministry of justice and CNRS. Our proposal is a joint project with L. Godefroy (Faculté de droit et science politique, Nice University), who has expertise in the relations between the digital world and law, and F. Lebaron (Versailles St Quentin University). F. Lebaron is a sociologist and a specialist of performativity. We aim at studying the performative effects of scales from a general point of view by using our respective knowledges in law, sociology and statistics. More precisely, we will first choose some domains where scales have been introduced, like for instance child support or competition law. Statistical studies based on sociological insights will then be performed to measure how much these scales have performed as compared to the previous, scale-free, situation. This step will require to construct models in order to enhance the estimation step and thus the interpretation of the results. Based on the analysis of the current performative effects and our models, we will, if needed, propose modifications allowing one to reduce unwanted effects.

As a last step, we hope that a global pattern of how scales perform will emerge, maybe from a comparative analysis of the models in different areas. This could open the way to the construction a general theory.

4.2.3. Quantifying legal risk

Our most successful application to date is in the quantification of legal risk: once one is prepared to accept that a legal decision is a random variable, one realizes that legal risk, which is a special component of the global risk companies or even citizens face, may be treated as are other risks. In particular, financial risks have been the topic of extensive studies in recent years, partly in response to the several crises we have witnessed. One lesson from this area is that, although one cannot of course predict the future state of a market, one is able to estimate its probability distribution. This allows one for instance to compute Values at Risk and thus to control one's risk.

We have designed an approach that can quantify legal risk in the same way as financial risk: given a specific domain, e.g. spousal support or dismissal without fair cause, we carefully design a set of legal criteria and analyse a large amount of cases in light of these criteria. We then use refined machine learning techniques to produce a probability distribution that reflects the decisions that would be taken by the judges in our database. This probability distribution takes into account both inter- and intra-judges variability. The mathematical result is that, when the size of the database tends to infinity, the estimated probability distribution tends, under some assumptions, to the actual one.

We have applied this theory to two fields so far : spousal support and dismissal without fair cause. Our future plans include in particular areas in labour law.

In view of to the strong interest this tool has raised among professionals (lawyers, insurance companies, but also the french ministry of justice), we are thinking of creating a start-up company that would commercialize it. As a consequence, we are not able to detail the mathematics involved in this study.

4.2.4. Intellectual property

This project is conducted in the frame of an ISN-funded collaboration between Inria and CERDI (University Paris Sud). Its aim is to help judges make informed decisions concerning the amount of fines in cases of violation of intellectual property. Indeed, in this domain, the fundamental rule that the amount is fixed so as to make good the damage suffered is not adequate: a person who commits a fault with a view to gain can be condemned, in addition to compensatory damages, to pay punitive damages. This rule has been introduced in 2007 under the impulsion of European law. In practice, it seems that it has not been implemented with great success. Our contribution studies a Bayesian network model for understanding how judges compute such amounts. We construct two such networks, one based on law and jurisprudence from Canada and one from France. This project has started in the fall of 2015.

4.3. Archaeology

We have been working since 2011 on the construction of new Bayesian approach for chronological modeling: this is an important issue in archaeology and paleo-environmental sciences. The archaeologists base their interpretations on a wide range of sources of information. A priori knowledge about the parameters of the model is often available, and so it should be considered along with the model and the data. This motivates the Bayesian choice.

In our case the data are the measurements M_i provided by dating laboratories e.g. 14C). The prior information contains historical evidence (e.g. an event must have occurred between two calendar dates,..) or geological information (e.g. a stratigraphic information,..). All the measurements require a calibration step to be converting into calendar date.

Tools for Constructing Chronologies

The aim is to provide probabilistic estimation of a chronology; a crucial aspect is to obtain a robust approach with respect to outliers due to the sampling in the field or the measurement process in the laboratory.

The solution proposed in [7], [6] is based on the "event model'. We define the Event as the date θ of an archeological context determined from a collection of contemporaneous artifacts. The model with random effect can be written as follows

$$\begin{aligned} M_i &= g_i(t_i) + S_i\rho_i \\ t_i &= \theta + \sigma_i\lambda_i \end{aligned}$$

where g_i is the calibration function and $(\rho_1, ..., \rho_n, \lambda_1, ..., \lambda_n)$ are iid standard Gaussian random variables. The random variables $(\lambda_i)_i$ and $(\epsilon_i)_i$ are interpreted as follows :

- $S_i\rho_i$ represents the experimental error provided by the laboratory and the calibration step.
- $\sigma_i\lambda_i$ represents the irreducible error between t_i and θ due to sampling problems external to the laboratory

In [7], [6], we show the ability of the variance σ_i^2 to take large values, in order to automatically penalize an outlier.

To enrich the chronological modelling, we wish to incorporate archaeological “phases”. Contrary to an “event”, a phase suggests duration. The objective is then to estimate the parameters that characterize the phase (beginning /end/duration), and then to develop Bayesian tests on the duration of the phase or the existence of a gap (hiatus) between two phases.

Calibration

The dating processes provide measurements, which are converted into calendar dates using calibration reference curves. We plan to explore issues related to calibration for different dating methods.

Optically stimulated luminescence (OSL) dating is a quantitative dating method to determine the time of last exposure of sand and silt to sunlight. Our aim is to complete the model constructed in [2] in order to obtain an OSL age determination.

We generally observe a overestimation of the age of a sample by OSL dating. This can be explaining by an insufficient resetting of the optically stimulated luminescence signal prior to sediment deposition. Therefore detection of so-called poor bleaching is of prime importance in OSL dating.

5. New Software and Platforms

5.1. New Software

5.1.1. Law

In order for our research to become used in a non-technological domain as is law, we make special efforts to develop user-friendly software tools. We have already made available an iOS application for computing the amount of child support. Another application is currently being developed for spousal support.

Future user-friendly software we intend to propose will be concerned with the areas of personal injury and intellectual property.

5.1.2. Economy and finance

The software toolbox FracLab (http://fraclab.saclay.inria.fr) contains a number of tools related to multistable processes, including their simulation and calibration. We intend to pursue these developments in particular by proposing methods dealing with self-stabilizing processes.

We also hope to be able to collaborate with financial institutions, in view of advancing the idea that current regulations, because they are model-based, perform the economy in a way that needs to be carefully scrutinized.

5.1.3. Archaeology

Our software ChronoModel provides a set of Bayesian statistical tools for constructing chronologies using measurement coming from different dating methods (e.g. 14C, TL/OSL, AM, typo-chronology). Chronomodel is a free and open-source cross-platform software that can be downloaded at http://www.chronomodel.fr

It offers a user-friendly interface for entering both data and archaeological facts : stratigraphy, temporal order constraints, contemporaneous events.

ChronoModel is currently used by researchers in laboratories of archaeology and earth science, and also by archaeologists working in the organizations of rescue archaeology.

6. New Results

6.1. Legal aspects of systems designed to judicial risk quantification

Participants : Jérôme Dupré

Within the ANJA team systems designed to calculate judicial risk using machine learning technology (AI) have been developed. A former French magistrate is one of the team member who has participated to these researches. In the meantime, he endeavored to contribute to design a legal framework applicable to this activity.

Artificial intelligence (AI), particularly when applied to justice, is liable to encounter rules of law, which are applicable even in the absence of a specific law. As with any new field of activity (eg the Internet), the notion of “legal vacuum” must not be confused with that of “legislative void”. It is therefore necessary to identify how to protect these technologies, what is the responsibilities of each, whether designer or / and user, that could already be applicable.

The two main concerns are relating to property and liability.

1.Regarding property, one can observe that predictive/quantitative solutions based on artificial intelligence result from a combination of technical criterion, databases, algorithms and software, each subject to specific legal protections. These elements may hence be protected by the copyright (for technical criterion); the database law, copyright and unfair competition (for databases); the trade secret (for algorithms), the copyright (for software). It may be questioned whether it would be irrelevant to create, ultimately, a unified legal status specific to this complex reality. But it is probably too early to legislate.

At the heart of the solution is the algorithm, an immaterial element which, in France, is the least well protected by law (it belongs to the domain of “ideas”), justifying its secret nature.

2. Considering liability aspects, we observe that this “black box” - the secret being a consequence of complexity and investments made - may be at the origin of a prejudice, either because of a bad use, or because it is not correctly designed.

The French law offers a range of solutions to the victim, depending on the origin of the damage (see 6.2).

Trust in results is also a factor to be considered. Thus, in the absence of a technical problem peculiar to the solution, misuse by the legal professional providing legal advice may justify, for example, his/her contractual liability. From this standpoint, the reliance granted to the technology and the way it is presented are essential. It justifies a specific attention to the way contracts relating to these services are drafted.

The designer of a defective solution may be required to guarantee against the hidden defects of Article 1641 of the French Civil Code. (When there is no contract, one can also be liable on the grounds of Article 1242 paragraph 1 of the same Code).

Standardization of algorithms, which could be tested by an independent body and subject to secrecy, is also an option, but presents a risk of possible paralysis of a promising market in the field of mathematics.

More generally, it seems necessary to comply with the CNIL (the French Data Protection Authority) provisions relating to personal data (Law No 78-17 of 6 January 1978, spec. article 10, and soon Regulation EU 2016/679 of the European Parliament and of the Council of 27 April 2016 applicable in 2018) and with Privacy Law... A huge amount of data is indeed likely to reveal information not resulting from each data taken separately. But this risk is probably more present in the field of big data than in algorithms, the data used for learning being "dissolved" in the formula.

6.2. Liability and ethics

Participants : Jacques Lévy Véhel, Jérôme Dupré

Some legal issues are specifically related to the calculation error. The system user may not have entered the data correctly and then will be responsible for the displayed result (but it is possible to display the entered data with the results to limit this risk).

The user may also have entered the data correctly and: -he/she is aware of the error of the system, in which case the user regains a share of responsibility, -he/she is not aware of the error of the system and the error is not easily detectable, then the responsibility should move towards the authors of the system conception (with its different steps that result from each other: definition of the search criteria, creation of database and creation of the algorithm, development and integration of the software, etc.). -he/she is not aware of the error of the system and the error is easily detectable, in which case the user should regain at least some responsibility.

The use of systems designed to judicial risk quantification entails paradigm shift The probabilistic approach, where all things are equal, seems very far from the hierarchy of legal norms and legal causality. In the absence of any upheaval in the law, we can at least expect an upheaval in legal practice.

Ultimately, the use of predictive tools, which favors the discovery of correlations, may lead to less attention to the causes of events, hence the need to maintain a vigilance on this aspect. The human who "delegates everything" to the machine should not avoid responsibility.

From an ethical perspective, is it acceptable to calculate the sentence which is likely to be pronounced in penal law, e.g. for a crime? Should we accept to profile judges using all their past rulings?

More generally, a reflection as well as a study on the place left to the human at all the stages of the process of elaboration and use (the decision-making) of the predictive tools seems necessary.

6.3. Statistical inference methods for panel of random-coefficient AR(1) data

Participants: A. Philippe and D. Surgailis R. Leipus and V. Pilipauskaitė (Vilnius university)

We study the statistical inference methods for panel of random-coefficient AR(1) data [17] . We propose a nonparametric estimation of the distribution function of random coefficients by the empirical distribution of lag 1 sample correlations of individual AR(1) processes. Consistency and asymptotic normality of the empirical distribution function and a class of kernel density estimators is established under some regularity conditions on G(x) as N and n increase to infinity. An extension of this work consists in testing the presence of long memory. The procedure is based on the tail index of G and the théorie of extreme values. In the same direction, a new frequency-domain test statistic is introduced to test for short memory versus long memory, see [23]

6.4. New Bayesian approach for chronological modeling

Participants: Anne. Philippe and Marie-Anne Vibet in collaboration with IRAMAT - Université Bordeaux Montaigne

We have been working on the construction of new Bayesian approach for chronological modeling: this is an important issue in archaeology and paleo-environmental sciences. The proposed solution is based on the "event model'. We define the Event as the date of an archeological context determined from a collection of contemporaneous artifacts. We obtain a robust approach with respect to outliers due to the sampling in the field or the measurement process in the laboratory.

In [25] We propose new tools to analyse the chronologies especially regarding phases. They are implemented in R package 'RChronoModel'.

In [22], [18], [5], we propose bayesian models for optically stimulated luminescence dating. It consists in estimating a central equivalent dose from a set of luminescence measurements. Then a calibration step is required to convert equivalent dose into calendar date.

6.5. Self-regulated processes

Participants : Jacques Lévy Véhel, Anne Philippe, Caroline Robet
We wish to construct various instances of processes Z such that, at each point t, almost surely, the pointwise Hölder exponent of Z at t, denoted $\alpha_Z(t)$, verifies

$$\alpha_Z(t) = g(Z(t))$$

where $g \in \mathcal{C}^1(\mathbb{R}, [a,b])$ is a deterministic function. Then, we would estimate the function g which control the regularity.

The pointwise Hölder exponent at t of a function or a process $f : \mathbb{R} \to \mathbb{R}$, which is $\mathcal{C}^1$ nowhere, is the real $\alpha_f(t)$ such that :

$$\alpha_f(t) = \sup\{\beta, \limsup_{h\to 0} \frac{| f(t+h) - f(t) |}{| h |^\beta} = 0\}$$

We worked first on pathwise integrals :

Theorem 1 Let $g \in \mathcal{C}^1(\mathbb{R}, [a,b])$, $0 < a < b < 1$. Provided $\|g'\|_\infty$ is small enough, there exists a unique continuous process Z verifying almost surely on $[0,T]$

$$Z_t = \int_0^t (t-u)^{g(Z_u)-1} W_u \, du$$

where W is an almost surely continuous process.

A random condition ($\|g'\|_\infty \|W(\omega)\|_\infty C(a,T) < 1$) appears in the application of Banach fixed point theorem (in $(\mathcal{C}^0([0,T];\mathbb{R}), \|.\|_\infty)$). It implies that it is possible to have existence et uniqueness only on $[0,t']$, $t' < T$. We simulated pathwise integrals and showed some cases without uniqueness. We studied some easier processes in order to find the regularity of Z.

Theorem 2 Let $h \in]0,1[$ and U defined on $[0,T]$ by

$$U_t = \int_0^t (t-u)^{h-1} W_u \, du$$

Then $\forall t \in [0,T]$, $\alpha_U(t) \geq h$.

Theorem 3 Let $g \in \mathcal{C}^1(\mathbb{R}, [a,b])$, $0 < a < b < 1$. Provided $\|g'\|_\infty$ is small enough, there exists a unique continuous process Y verifying almost surely on $[0,T]$

$$Y_t = \int_0^t (t-u)^{g(Y_t)-1} W_u \, du$$

where W is an almost surely continuous process. Furthermore, $\forall t \in [0,T]$, $\alpha_Y(t) \geq g(Y_t)$

Then, we adapted the multifractional Brownian Motion [50], [31] (which a representation is $B_t = \int_0^t K_{H(t)}(t,u)W(du)$, W Brownian Motion et H$\in \mathcal{C}^1$) to construct the modified multifractional Brownian Motion : $Z_t = \int_0^t K_{H(u)}(t,u)W(du)$. We expect obtain a self-regulated process $Y_t = \int_0^t K_{g(Y_u)}(t,u)dW(u)$.

Theorem 4 Let $g \in \mathcal{C}^1(\mathbb{R},[a,b])$, $0 < a < b < 1$. Provided $\|g'\|_\infty$ is small enough, there exists a unique continuous adapted process Y include in $\mathcal{C}^0([0,T];L^2(\Omega))$ verifying almost surely on $[0,T]$

$$Y_t = \int_0^t K_{g(Y_u)}(t,u)dW(u)$$

where W is the Brownian motion.

6.6. Causal inference by independent component analysis with Application of to American macro-economic data

Participants : Jacques Lévy-Vehel, Anne Philippe, Marie-Anne Vibet

The aim of this work is to study the causal relationships existing among macro-economic variables under investigation, and trace out how economically interpreted random shocks affect the system. Structural vector of autoregressive models (SVAR) are usually applied in this kind of study and the causal structure is driven by the data. In this work, independent component analysis (ICA) is implemented in order to guaranty the identifiability of the causal structure. However, the use of ICA can only be done under the hypothesis that the residuals are non-Gaussian, an hypothesis easily verified with economic data.

The vector of autoregressive (VAR) model has the following reduced representation :

$$Y_t = A_1 Y_{t-1} + ... + A_p Y_{t-p} + u_t \text{ , for } t = 1, ..., T$$

where, Y_t, is the vector of contemporaneous variables of dimension $K \times 1$, p is the number of autoregressive variables, A_j, for $j = 1, ..., p$, are matrices of dimension $K \times K$ estimated by the model and u_t is the vector of random disturbances of dimension $K \times 1$ and assumed to be a zero-mean white noise process, $u_t \sim N(0_K, \Sigma_u)$. Given enough data, both Σ_u and all matrices A_j can be corretly estimated by the VAR model.

However, the VAR model is not sufficient for policy analysis. Indeed, using the Moving Average representation of a stable VAR :

$$Y_t = \sum_{j=0}^{\infty} \Phi_j u_{t-j} \tag{1}$$

where $\Phi_0 = I_K$ and $\Phi_j, j \geq 1$, are the coefficients matrices representing the impulse responses of the elements of Y_t to the disturbances u_{t-j}. This representation is not unique.

The structiral VAR (SVAR) is essentially a VAR equipped with a particular choice of a matrix P so that $Y_t = \sum_{j=0}^{\infty} \Phi_j P P^{-1} u_{t-j} = \sum_{j=0}^{\infty} \Psi_j \epsilon_{t-j}$

where ϵ_{t-j} are independent random shocks economically interpreted. To this aim, the ICA procedure is then used to find the proper matrix P using the hypothesis that the residuals, ϵ_{t-j}, are non-Gaussian.

We used the VAR-LINGAM procedure developped by Moneta *et al* [28] and their package written for R software. We started by testing this procedure with a series of simulations study. We tackled the following questions : Are the coefficients of the matrices B and A well estimated by the VARLINGAM procedure ? Is the bootstrap function appropriate, and in particular, does it estimate properly the standard error of the coefficients of matrices A and B ? And how long should the economic data be in order to estimate correctly the coefficients of the matrices B ?

As the conclusion to all these studies were correct enough, we went on analysing our real data that consists of 6 weekly time series US macro-economic data, reported from the first week of January 1996 to April 2016 : The BofA Merrill Lynch US Broad Market Index, The Bofa Merril Lynch US Corporate Index, Equity Indices S&P, 500, Federal Funds Rates,Treasury Bills, Other Factors Draining Reserve Balances.
The conclusions of this work is in discussion with economists and a paper will soon be written.

6.7. SAR image denoising using an irregularity-preserved denoising technique based on the global Höllder exponent

Participants : Jacques Lévy-Vehel, Yue Huang

This work addresses the speckle noise reduction for SAR images by using the irregularity-preserved denoising technique proposed in [34]. This irregularity preserving denoising scheme in [34] may be summarized as a three-step process in the following:

1. Apply a Discrete Wavelet Transform (DWT) on the noisy signal and represent the resulting coefficients distribution over scales. Estimate the cut-off scale and the global Hölder exponent α_f using linear regression of $\max_k \left(\log_2 | \langle f, \psi_{j,k} \rangle | \right)$ at larger scales.
2. Extrapolate the larger scale regression line to smaller scales and limit coefficient at smaller scales ($j \geq j_{\text{cut-off}}$) to the boundary value obtained from the linear regression
3. Reconstruct the filtered signal from the set of modified coefficients

where f is the signal under analysis, $\psi_{j,k}$ is the wavelet basis, and $\langle f, \psi_{j,k} \rangle$ is the wavelet coefficient of f at scale j and location k. As it has been shown by simulations in [34], to retrieve irregular signals affected by additive noise, this technique outperforms conventional denoising techniques that apply hard or soft thresholding to the wavelet coefficients.

Considering a speckle-affected SAR image, a complex SAR signal may be represented by:

$$y(l) = s(l)u(l)$$

where l represents one of L realizations, and the noise term $u(l)$ follows a complex circular centered Gaussian white distribution with unit variance, i.e. $u \sim \mathcal{N}_C(0,1)$, $E(u(i)u^*(j)) = \delta_{(i-j)}$. The texture of SAR image significantly depends on the backscattering power $\sigma(l) = |y(l)|^2$.

We aim to use the irregularity-preserved denoising technique to denoise SAR image and enhance its texture. We tested firstly on the simulated signals affected by multiplicative noise and then on real SAR images. This denoising scheme showed potential to reduce the speckle noise, preserve the irregularity of image texture and enhance target signature.

Although the results have been compared with other SAR speckle filtering techniques, we still need more efforts for validation. As long as the results are validated, the work will be written in a paper.

6.8. Underfoliage object imaging using SAR tomography and wavelet-based sparse estimation methods

Participants : Yue Huang, Jacques Lévy-Vehel

Hybrid environments refer to a scenario of deterministic objects embedded in a host natural random environment and their scattering patterns consist of a complex mixture of diverse mechanisms, like, in the case of this study, volume scattering from the canopy, double bounce reflection between the ground and under-foliage objects as well as between objects and trunks, surface scattering from the underlying ground, etc. The resulting SAR information is characterized by a strong complexity, and its analysis using 2-D images or even data acquired in InSAR configuration remains difficult. Using Multi-baseline(MB) InSAR data, SAR tomography can be applied to reconstruct in 3-D the measured scattering responses and polarimetric patterns. Natural volumes, such as forest canopies, being composed of a large number of scatterers whose responses cannot be discriminated at the resolution of analysis, their scattering patterns are generally considered as a vertical density of random or speckle-affected reflectivity. On the other hand, localized objects, such as artificial targets on the ground are associated to point-like contributions, that may be separable in the vertical direction. The global response of under-foliage objects with a deterministic scattering response embedded in surrounding distributed environments, can be described by a mixed spectrum. Conventional tomographic techniques like the Capon and Beamforming methods, estimate continuous Power Spectral Density (PSD) and hence are well adapted to the characterization of continuous volumetric media, but cannot discriminate closely-spaced scatterers, e.g. scattering responses from trucks, due to limited spatial resolution. Conventional high-resolution methods like MUSIC and subspace fitting estimators as well as sparse estimation techniques such as LASSO [52] and FOCUSS [40], are well adapted to the characterization of discrete scatterers like truck top, truck-ground interaction and calibrators over bare soils, or buildings over urban areas [53], but cannot properly handle the high dimensionality of the scattering responses of natural volumes. Usual tomographic techniques cannot simultaneously cope with both types of spectrum, and not able to deal with mixed spectral estimation problems, characteristic of underfoliage object imaging scenario.

Wavelet-based techniques present a high potential for such applications, as they permit to parameterize in a sparse way continuous functions, i.e. canopy PSDs in the present case. Wavelet-based tomographic techniques have been used for tomographic imaging of forested areas [27], and for such regular signals, large wavelet coefficients being often concentrated in the approximation space, scale thresholding may be implemented to extract the most significant wavelet coefficients for an accurate volume signal recovery [27]. In the underfoliage object scenario, discrete scatterers embedded in a continuous medium, result in a mixed vertical PSD that may be associated to an irregular signal with wavelet coefficients distributed both in the approximation and detail spaces, and a simple scale cut-off is hence not adapted to separate the wavelet coefficients of discrete scatterers from those of continuous media. Therefore, we propose a new wavelet-based method to extract underfoliage objects from their speckle-affected distributed environment and characterize them with a high resolution.

For an MB-InSAR configuration with M acquisition positions, considering an azimuth-range resolution cell containing a mixture of backscattering contributions from object (o) and volume (v) scatterers located at different heights z, the observed data vector at lth realization can be represented by:

$$\mathbf{y}(l) = \mathbf{A}_o(\mathbf{z}_o)\mathbf{s}_o(l) + \mathbf{A}_v(\mathbf{z}_v)\mathbf{s}_v(l) + \mathbf{n}(l) \tag{2}$$

where the steering matrix, $\mathbf{A}_x(\mathbf{z}_x)$, contains the interferometric phase information associated to the InSAR responses of the scatterers located at the unknown elevation positions $\mathbf{z}_x = [z_{x_1}, \cdots, z_{x_{N_x}}]$ above the reference focusing plane, and the source signal vector, $\mathbf{s}_x = [s_{x_1} \cdots s_{x_{N_x}}]^T \in \mathbb{C}^{N_x \times 1}$, contains the unknown complex backscattering coefficients of the N_x source scatterers. The vertical reflectivity function can be represented as $\mathbf{p}_x = E(|\mathbf{s}_x|^2)$ $(x = o, v)$.

Over speckle-affected environments, unknown reflectivity and elevation parameters are generally estimated from second-order statistics, i.e. from the covariance matrix $\widehat{\mathbf{R}} \in \mathbb{C}^{M \times M}$ of the observed MB-InSAR data $\mathbf{y} \in \mathbb{C}^{M \times 1}$. The proposed tomographic processing technique is based on the minimization of the Least-Square (LS) fitting between the observed and modeled data covariance $||\mathbf{R} - \widehat{\mathbf{R}}||_F$. The modeled covariance matrix is composed by the covariances of object and volume contributions $\mathbf{R} = \mathbf{R}_o + \mathbf{R}_v$, each of them being simply related to its discretized vertical density of reflectivity $\mathbf{p}_x$ through $\mathbf{R}_x = \mathbf{A}(\mathbf{z}_x)\,\mathrm{diag}\,(\mathbf{p}_x)\mathbf{A}^H(\mathbf{z}_x) \in \mathbb{C}^{M \times M}$.

The proposed method can be represented by a l_1 norm minimization in a transformed space subject to quadratic constraints between the observed and modeled data covariance:

$$\min_{\mathbf{p}} ||\mathbf{Bp}||_1 \text{ subject to } \quad ||\mathbf{R} - \widehat{\mathbf{R}}||_F \leq \epsilon \tag{3}$$

where

- $\mathbf{p} = [\mathbf{p}_o^T \quad \mathbf{p}_v^T]^T \in \mathbb{R}^{+N_s \times 1}$ stands for vertical backscattering power distribution for the resolution cell under analysis,
- $\mathbf{B} = \begin{bmatrix} \mathbf{I}_{(N_o \times N_o)} & \mathbf{0} \\ \mathbf{0} & \Psi_{(N_v \times N_v)} \end{bmatrix} \in \mathbb{R}^{(N_s \times N_s)}$ represents the hybrid sparsifying basis with the wavelet basis Ψ

This tomographic technique is suitable for the mixed-spectrum estimation problem, because it maintains the spectral continuity for the backscattering power of forest canopies and the high-resolution for the vertical reflectivity of objects. The effectiveness of this new approach is demonstrated using L-band airborne tomographic SAR data aquired by the DLR over Dornstetten, Germany. The undeniable performance can be shown by the results in [21] and [20].

This work has been presented in European SAR conference 2016 . Some refined results have been presented in IGARSS conference 2016 as an invited talk. By extending this work in details, a journal paper [24] has been submitted to IEEE Geoscience and Remote Sensing Letters (GRSL) and is currently under reviewing.

6.9. Detection of objects concealed beneath forest canopies using Time-Frequency techniques

Participants : Yue Huang, Jacques Lévy-Vehel

In the scenario of hybrid environments where objects with a deterministic response are embedded in a speckle affected environment, the parameter estimation for this type of scatterers becomes a problem of mixed-spectrum estimation. To isolate and characterize these different scattering contributions, a novel method proposed by Huang et al. was used to extract isolated scatterers (IS) from their surrounding distributed environments, named IS extraction in [42]. Incorporating the Weighted Subspace Fitting (WSF) estimator, this method estimated scattering responses within one resolution cell and then distinguishes isolated scatterers from distributed ones by calculating the cross-correlation between the measured data and the estimated scattering responses. Moreover, to compare the detection performance for coherent scatterers, two statistical methods have been applied to analyse hybrid environments in [43]: GLRT (generalized likelihood ratio test)-based and SSF (weighted Signal Subspace Fitting)-based detection procedures. However, the above mentioned methods based on discrete high-resolution tomographic estimation, require to preselect the number of scattering contributions, which may induce reliability issues due to model order selection.

This paper proposes a new tomographic estimator based on Time-Frequency (TF) techniques using Multi-baseline Polarimetric and Interferometric SAR data. The coherent TF analysis of polarimetric SAR has been introduced in [38], [39] for the study of anisotropic scattering behaviors and then applied in [37], [36] for dense urban environment characterization. Time-frequency techniques can represent spectral properties around specific spatial locations or spatial features at specific spectral positions, leading to describe local variations of spectral or spatial features. Considering SLC SAR images, the spectral locations can be linked to azimuth looking angle and illumination frequency in such a way:

$$w_{az} = \frac{4\pi}{c} f_c v_{SAR} \sin\phi, \quad w_{rg} = \frac{4\pi}{c}(f - f_c)$$

with f_c central frequency and ϕ azimuth looking angle. The TF technique can be used to analyze scattering behaviors at different illuminated positions and frequency components during SAR integration. Based on the correlation between different spectral positions, the TF indicator proposed in [37] can extract coherent components in complex random SAR responses. Polarimetric TF indicator has been developed in [41] for ship discrimination. In this paper, the new tomographic estimator extends 2-D TF analysis to 3-D, which provides an efficient cancellation for clutters from speckle-affected random scattering environments, and discriminates the deterministic responses from coherent scatterers in 3-D space. The effectiveness of this new tomographic approach is demonstrated by using L-band MB-PolInSAR data set acquired over the test site of Dornstetten where the underfoliage objects are set up. The fully polarimetric version of this TF tomographic estimator is also developed to improve the detection efficiency. This work has been accepted for oral presentation at the Polinsar 2017 Workshop and the final paper will be written by the end of Workshop.

7. Dissemination

7.1. Promoting Scientific Activities

7.1.1. Scientific Events Organisation

7.1.1.1. Member of the Organizing Committees

Bayesian workshop : Bayesian statistics applied to archaeology - May 2016 Nantes
Anne Philippe and Marie-Anne Vibet were main organizers.

7.1.2. Journal

Anne Philippe is an Associated editor of Computational Statistics and Data analysis
Jacques Lévy Véhel is associate editor of the journal « Fractals »

7.1.3. Invited Talks

Anne Philippe and Marie-Anne Vibet were invited to the Ibercrono conference, Barcelone Spain, October 2016
Anne Philippe was invited to ArcheoFoss, Cagliari Italy October 2016
Yue Huang gave a talk at the invited session of advanced SAR technologies at IEEE International Geoscience and Remote Sensing Symposium, Beijing Chine, 2016

7.1.4. Research Administration

Anne Philippe is a member of commits national du CNRS

7.2. Teaching - Supervision - Juries

7.2.1. Teaching

- Marie-Anne Vibet: Introduction to SAS software, Master 2 Ingénierie Mathématique, January and November 2016 (12hrs)
- Anne Philippe : Statistical inference, Master 1 Ingénierie Mathématique, Automne 2016 (24hrs)
- Anne Philippe : Bayesian statistics, Master 2 Ingénierie Mathématique, Automne 2016 (36hrs)

7.2.2. Supervision

- PhD in progress: Vytauté Pilipauskaité, supervised by Anne Philippe and Donatas Surgailis
- PhD in progress: Caroline Robet, supervised by Anne Philippe and Jacques Lévy-Véhel

7.2.3. Juries

Anne Philippe was in the jury of PhD theses :

- President of jury for Kaniav Canary
- Reporter and member of jury for Lilliam Urrego, Le Quyen Thieu

8. Bibliography

Major publications by the team in recent years

[1] O. BARRIERE, A. ECHELARD, J. LÉVY VÉHEL.*Self-Regulating Processes*, in "Electronic Journal of Probability", December 2012 [*DOI :* 10.1214/EJP.v17-2010], https://hal.inria.fr/hal-00749742.

[2] B. COMBÈS, A. PHILIPPE, P. LANOS, N. MERCIER, C. TRIBOLO, G. GUERIN, P. GUIBERT, C. LAHAYE.*A Bayesian central equivalent dose model for optically stimulated luminescence dating*, in "Quaternary Geochronology", 2015, vol. 28, p. 62 - 70 [*DOI :* 10.1016/J.QUAGEO.2015.04.001], http://www.sciencedirect.com/science/article/pii/S1871101415300029.

[3] A. ECHELARD, J. LÉVY VÉHEL, A. PHILIPPE.*Statistical estimation of a class of self-regulating processes*, August 2014, To appear in Scandinavian Journal of Statistics, https://hal.inria.fr/hal-00868604.

[4] K. J. FALCONER, J. LÉVY VÉHEL.*Multifractional, multistable, and other processes with prescribed local form*, in "Journal of Theoretical Probability", 2009, vol. 22, n^{o} 2, p. 375-401, The original publication is available at : https://www.springerlink.com [*DOI :* 10.1007/S10959-008-0147-9], https://hal.inria.fr/inria-00539033.

[5] M. FROUIN, S. HUOT, S. KREUTZER, C. LAHAYE, M. LAMOTHE, A. PHILIPPE, N. MERCIER.*An improved radiofluorescence single-aliquot regenerative dose protocol for K-feldspars*, in "Quaternary Geochronology", 2017, vol. 38, p. 13 - 24.

[6] P. LANOS, A. PHILIPPE.*Event model: a robust Bayesian tool for chronological modeling*, December 2015, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01241720.

[7] P. LANOS, A. PHILIPPE.*Hierarchical Bayesian modeling for combining Dates in archaeological context*, December 2015, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01162404.

[8] R. LE GUÉVEL, J. LÉVY VÉHEL.*A Ferguson - Klass - LePage series representation of multistable multifractional processes and related processes*, in "Bernoulli", 2012, vol. 18, n^{o} 4, p. 1099-1127 [*DOI :* 10.3150/11-BEJ372], https://hal.inria.fr/inria-00538985.

[9] R. LEIPUS, A. PHILIPPE, V. PILIPAUSKAITĖ, D. SURGAILIS.*Nonparametric estimation of the distribution of the autoregressive coefficient from panel random-coefficient AR(1) data*, in "Journal of Multivariate Analysis", January 2017, vol. 153, p. 121 - 135 [*DOI :* 10.1016/J.JMVA.2016.09.007], https://hal.archives-ouvertes.fr/hal-01174833.

[10] J. LÉVY VÉHEL, P.-E. LÉVY VÉHEL, V. LÉVY VÉHEL.*A model investigating incentives for illegal cultural goods sharing sites to become legal*, January 2015, working paper or preprint, https://hal.inria.fr/hal-01103048.

[11] J. LÉVY VÉHEL.*A simple isochore model evidencing regulation risk*, in "IME'2015, Liverpool", 2015.

[12] J. LÉVY VÉHEL, C. WALTER.*Prudential regulation: is there a danger in reducing the volatility?*, in "9-th Actuarial Research Conference, Santa Barbara", 2014.

[13] A. Philippe, M.-A. Vibet.*Analysis of archaeological phases using the CRAN package RChronoModel' A tutorial*, July 2016, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01347895.

Publications of the year

Articles in International Peer-Reviewed Journal

[14] F. Barsotti, A. Philippe, P. Rochet.*Hypothesis testing for Markovian models with random time observations*, in "Journal of Statistical Planning and Inference", June 2016, vol. 173, p. 87 - 98 [*DOI :* 10.1016/j.jspi.2015.12.009], https://hal.archives-ouvertes.fr/hal-01428011.

[15] L. J. Fermin, J. Lévy Véhel.*Variability and singularity arising from poor compliance in a pharmacokinetic model II: the multi-oral case*, in "Journal of Mathematical Biology", 2016 [*DOI :* 10.1007/s00285-016-1041-1], https://hal.inria.fr/hal-00868621.

[16] M. Frouin, S. Huot, S. Kreutzer, C. Lahaye, M. Lamothe, A. Philippe, N. Mercier.*An improved radiofluorescence single-aliquot regenerative dose protocol for K-feldspars*, in "Quaternary Geochronology", March 2017, vol. 38, p. 13 - 24 [*DOI :* 10.1016/j.quageo.2016.11.004], https://hal.archives-ouvertes.fr/hal-01427990.

[17] R. Leipus, A. Philippe, V. Pilipauskaitė, D. Surgailis.*Nonparametric estimation of the distribution of the autoregressive coefficient from panel random-coefficient AR(1) data*, in "Journal of Multivariate Analysis", January 2017, vol. 153, p. 121 - 135 [*DOI :* 10.1016/j.jmva.2016.09.007], https://hal.archives-ouvertes.fr/hal-01174833.

[18] N. Mercier, S. Kreutzer, C. Christophe, P. Guibert, C. Lahaye, P. Lanos, A. Philippe, C. Tribolo.*Bayesian statistics in luminescence dating: The 'baSAR'-model and its implementation in the R package 'Luminescence'*, in "Ancient TL", 2016, vol. 34, n^o 2, p. 14-21, https://hal.archives-ouvertes.fr/hal-01451482.

[19] R. Stoica, A. Philippe, P. Gregori, J. Mateu.*ABC Shadow algorithm: a tool for statistical analysis of spatial patterns stat*, in "Statistics and Computing", 2017 [*DOI :* 10.1007/s11222-016-9682-x], https://hal.archives-ouvertes.fr/hal-01427935.

Invited Conferences

[20] Y. Huang, J. Levy Vehel, L. Ferro-Famil, A. Reigber.*3D imaging for underfoliage targets using L-band multibaseline polinsar data and sparse estimation methods*, in "IGARSS 2016 - IEEE International Geoscience and Remote Sensing Symposium", Beijing, China, July 2016, https://hal.inria.fr/hal-01418435.

International Conferences with Proceedings

[21] Y. Huang, J. Lévy Véhel, L. Ferro-Famil, A. Reigber, S. Fortunati.*3D Characterization of Underfoliage Targets Using L-band To- mographic SAR Data and A Wavelet-Based Approach*, in "EUSAR 2016 - 11th European Conference on Synthetic Aperture Radar", Hamburg, Germany, June 2016, https://hal.inria.fr/hal-01418677.

Other Publications

[22] B. COMBÈS, A. PHILIPPE.*Bayesian analysis of individual and systematic multiplicative errors for estimating ages with stratigraphic constraints in optically stimulated luminescence dating*, November 2016, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01346868.

[23] G. GROMYKOV, M. OULD HAYE, A. PHILIPPE.*A Frequency-domain test for long range dependence*, August 2016, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01353483.

[24] Y. HUANG, J.-J. LEVY, L. FERRO-FAMIL, A. REIGBER.*3D Imaging Of Objects Concealed Below A Forest Canopy Using SAR Tomography At L-Band Data And Wavelet-Based Sparse Estimation*, January 2017, working paper or preprint, https://hal.inria.fr/hal-01451241.

[25] A. PHILIPPE, M.-A. VIBET.*Analysis of archaeological phases using the CRAN package RChronoModel' A tutorial*, July 2016, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01347895.

References in notes

[26] *Volatility stirs, markets unshaken*, in "BIS Quarterly Review, BIS Monetary and Economic Department", 2014.

[27] E. AGUILERA, M. NANNINI, A. REIGBER.*Wavelet-Based Compressed Sensing for SAR Tomography of Forested Areas*, in "Geoscience and Remote Sensing, IEEE Transactions on", Dec 2013, vol. 51, n^{o} 12, p. 5283-5295, http://dx.doi.org/10.1109/TGRS.2012.2231081.

[28] M. ALESSIO, E. DORIS, H. PATRIK, C. ALEX.*Causal Inference by Independent Component Analysis with Applications to Micro- and Macroeconomic Data*, in "Jena Economic Research Papers", 2010.

[29] S. BIANCHI.*Pathwise Identification Of The Memory Function Of Multifractional Brownian Motion With Application To Finance*, in "International Journal of Theoretical and Applied Finance", June 2005, https://ssrn.com/abstract=1888192.

[30] C. BORIO.*interview in Die Welt*, october 2014.

[31] B. BOUFOUSSI, M. DOZZI, R. MARTY.*Local time and Tanaka formula for a Volterra-type multifractional Gaussian process*, in "Bernoulli", 2010, https://hal.archives-ouvertes.fr/hal-00389740.

[32] P. CARR, H. GEMAN, D. B. MADAN, M. YOR.*The Fine Structure of Asset Returns: An Empirical Investigation*, 2000.

[33] S. DEMIRALP, K. HOOVER.*Searching for the Causal Structure of a Vector Autoregression*, in "Oxford Bulletin of Economics and Statistics 65, Supplement, 0305-9049", 2003.

[34] A. ECHELARD, J. LÉVY VÉHEL.*Local Regularity Preserving Signal Denoising I: Hölder Exponents*, November 2013, submitted, https://hal.inria.fr/hal-00879754.

[35] F. FERRARO, J. PFEFFER, R. I. SUTTON.*Economics Language and Assumptions: How Theories Can Become Self-Fulfilling*, in "The Academy of Management Review", 2005, vol. 30, n^{o} 1, p. 8-24, http://www.jstor.org/stable/20159091.

[36] L. FERRO-FAMIL, M. NEUMANN.*Recent advances in the derivation of Pol-InSAR statistics: study and applications*, in "Proc. EUSAR", June 2008.

[37] L. FERRO-FAMIL, E. POTTIER.*Urban area remote sensing from L-band PolSAR data using Time-Frequency techniques*, in "IEEE Urban Remote Sensing Joint Event", April 2007, p. 5045-5048.

[38] L. FERRO-FAMIL, A. REIGBER, E. POTTIER, W. BOERNER.*Scene characterization using sub-aperture polarimetric SAR data analysis*, in "IEEE Trans. Geosc. and Remote Sensing", Oct. 2003, vol. 41, n^o 10, p. 2264-2276.

[39] L. FERRO-FAMIL, A. REIGBER, E. POTTIER.*Nonstationary natural media analysis from polarimetric SAR data using a two-dimensional time–frequency decomposition approach*, in "Can. J. Remote Sensing", 2005, vol. 31, n^o 1, p. 21–29.

[40] I. F. GORODNITSKY, J. S. GEORGE, B. D. RAO.*Neuromagnetic source imaging with FOCUSS: A recursive weighted minimum norm algorithm*, in "J. Electroencephalog. Clinical Neurophysiol", Oct 1995, vol. 95, n^o 4, p. 231-251.

[41] C. HU, L. FERRO-FAMIL, G. KUANG.*Ship Discrimination Using Polarimetric SAR Data and Coherent Time-Frequency Analysis*, in "Remote Sensing", 2013, vol. 5, n^o 12, p. 6899–6920 [*DOI :* 10.3390/RS5126899], https://hal.archives-ouvertes.fr/hal-01128573.

[42] Y. HUANG, L. FERRO-FAMIL, C. LARDEUX.*Natural environment characterization using hybrid tomographic approaches*, in "Proc. PolInSAR", Jan 2011.

[43] Y. HUANG, L. FERRO-FAMIL, A. REIGBER.*Under-foliage target detection using multi-baseline L-band PolInSAR data*, in "Radar Symposium (IRS), 2013 14th International", Dresden, Germany, June 2013, https://hal.archives-ouvertes.fr/hal-01142799.

[44] D. JON, J.-P. ZIGRAND, H. S. SHIN.*Asset Price Dynamics with Value-at-Risk Constrained Traders*, June 2001, http://dx.doi.org/10.2139/ssrn.302307.

[45] R. LE GUÉVEL.*An estimation of the stability and the localisability functions of multistable processes*, in "Electronic journal of statistics ", 2013, vol. 7, p. 1129-1166 [*DOI :* 10.1214/13-EJS797], https://hal.archives-ouvertes.fr/hal-00480881.

[46] D. MACKENZIE, F. MUNIESA, L. SIU.*Do Economists Make Markets?*, Princeton University Press, 2007.

[47] J. MCGINNIS, S. WASICK.*Law's Algorithm*, 2014, http://ssrn.com/abstract=2130085.

[48] F. MUNIESA, M. CALLON.*La performativité des sciences économiques*, 2008, CSI WORKING PAPERS SERIES 010, https://halshs.archives-ouvertes.fr/halshs-00258130.

[49] J. PEARL, M. GLYMOUR, N. P. JEWELL.*Causal Inference in Statistics: A Primer*, in "Wiley", 2016.

[50] R.-F. PELTIER, J. LÉVY VÉHEL.*Multifractional Brownian Motion : Definition and Preliminary Results*, Inria, 1995, n^o RR-2645, Projet FRACTALES, https://hal.inria.fr/inria-00074045.

[51] Q. PENG.*Multifractional Stochastic Volatility Models: Holder regularity, statistical inference and simulation*, in "Int. Conf. Sto. Proc. App.", 2009.

[52] R. TIBSHIRANI.*Regression Shrinkage and Selection Via the Lasso*, in "Journal of the Royal Statistical Society, Series B", 1994, vol. 58, p. 267-288.

[53] X. X. ZHU, R. BAMLER.*Tomographic SAR Inversion by L1-Norm Regularization-The Compressive Sensing Approach*, in "IEEE Transactions on Geoscience and Remote Sensing", Oct 2010, vol. 48, n^{o} 10, p. 3839-3846.

Project-Team ASAP

As Scalable As Possible: Foundations of large scale dynamic distributed systems

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

IN PARTNERSHIP WITH:
Institut national des sciences appliquées de Rennes
Université Rennes 1

RESEARCH CENTER
Rennes - Bretagne-Atlantique

THEME
Distributed Systems and middleware

Table of contents

Project-Team ASAP

Creation of the Project-Team: 2007 July 01

Keywords:

Computer Science and Digital Science:

1.1.1. - Multicore
1.1.6. - Cloud
1.1.7. - Peer to peer
1.1.9. - Fault tolerant systems
1.2.9. - Social Networks
1.3. - Distributed Systems
1.5.2. - Communicating systems
2.1.6. - Concurrent programming
2.1.7. - Distributed programming
2.6.2. - Middleware
3.1.3. - Distributed data
3.1.8. - Big data (production, storage, transfer)
3.5.1. - Analysis of large graphs
3.5.2. - Recommendation systems
4.8. - Privacy-enhancing technologies
7.1. - Parallel and distributed algorithms
7.9. - Graph theory

Other Research Topics and Application Domains:

6.3.1. - Web
6.3.3. - Network Management
6.3.4. - Social Networks
6.4. - Internet of things
6.5. - Information systems
9.4.1. - Computer science
9.8. - Privacy

1. Members

Research Scientists

Anne-Marie Kermarrec [Inria, Senior Researcher, **on secondment at Mediego since Apr 2015**, HDR]
Davide Frey [Inria, Researcher]
George Giakkoupis [Inria, Researcher]

Faculty Members

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

Technical Staff

Alexandre Van Kempen [Inria, from Mar 2016]

PhD Students

Fabien Andre [Technicolor, until Nov 2016, granted by CIFRE]
Simon Bouget [Univ. Rennes I]
Adrien Luxey [Univ. Rennes I, from Oct 2016]
Nupur Mittal [Inria, until Nov 2016, granted by Conseil Régional de Bretagne]
Javier Olivares [Inria, until Dec 2016, granted by Google Inc]
Antoine Rault [Inria, until Jun 2016]
Pierre-Louis Roman [Univ. Rennes I]
Olivier Ruas [Inria, granted by Google Inc]
Resmi Ariyattu Chandrasekharannair [Univ. Rennes I, granted by UEB]
Stephane Delbruel [Univ. Rennes I]

Post-Doctoral Fellows

Zohir Bouzid [Univ. Rennes I]
Hicham Lakhlef [Univ. Rennes I, until Aug 2016]

Visiting Scientist

Arvid Bosk [KTH Royal Institute of Technology, Guest PhD Student, from Dec 2016]

Administrative Assistant

Virginie Desroches [Inria]

Others

Julien Lepiller [Inria, Student, from Feb to Jun 2016]
Florestan de Moor [Univ. Rennes I, Student, from May 2016 until July 2016]

2. Overall Objectives

2.1. Decentralized personalization

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

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

2.2. Scalability: Cloud computing meets p2p

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

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

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

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

2.3. Privacy-aware decentralized computations

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

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

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

2.4. Information dissemination over social networks

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

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

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

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

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

2.5. Computability and efficiency of distributed systems

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

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

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

3. Research Program

3.1. Theory of distributed systems

Finding models for distributed computations prone to asynchrony and failures has received a lot of attention. A lot of research in this domain focuses on what can be computed in such models, and, when a problem can be solved, what are its best solutions in terms of relevant cost criteria. An important part of that research is focused on distributed computability: what can be computed when failure detectors are combined with conditions on process input values for example. Another part is devoted to model equivalence. What can be computed with a given class of failure detectors? Which synchronization primitives is a given failure class equivalent to? These are among the main topics addressed in the leading distributed computing community. A second fundamental issue related to distributed models is the definition of appropriate models suited to dynamic systems. Up to now, the researchers in that area consider that nodes can enter and leave the system, but do not provide a simple characterization, based on properties of computation instead of description of possible behaviors [58], [51], [53]. This shows that finding dynamic distributed computing models is today a "Holy Grail", whose discovery would allow a better understanding of the essential nature of dynamic systems.

3.2. Peer-to-peer overlay networks

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

A peer-to-peer overlay network logically connects peers on top of IP. Two main classes of such overlays dominate, structured and unstructured. The differences relate to the choice of the neighbors in the overlay, and the presence of an underlying naming structure. Overlay networks represent the main approach to build large-scale distributed systems that we retained. An overlay network forms a logical structure connecting participating entities on top of the physical network, be it IP or a wireless network. Such an overlay might form a structured overlay network [59], [60], [61] following a specific topology or an unstructured network [56], [62] where participating entities are connected in a random or pseudo-random fashion. In between, lie weakly structured peer-to-peer overlays where nodes are linked depending on a proximity measure providing more flexibility than structured overlays and better performance than fully unstructured ones. Proximity-aware overlays connect participating entities so that they are connected to close neighbors according to a given proximity metric reflecting some degree of affinity (computation, interest, etc.) between peers. We extensively use this approach to provide algorithmic foundations of large-scale dynamic systems.

3.3. Epidemic protocols

Epidemic algorithms, also called gossip-based algorithms [55], [54], constitute a fundamental topic in our research. In the context of distributed systems, epidemic protocols are mainly used to create overlay networks and to ensure a reliable information dissemination in a large-scale distributed system. The principle underlying technique, in analogy with the spread of a rumor among humans via gossiping, is that participating entities continuously exchange information about the system in order to spread it gradually and reliably. Epidemic algorithms have proved efficient to build and maintain large-scale distributed systems in the context of many applications such as broadcasting [54], monitoring, resource management, search, and more generally in building unstructured peer-to-peer networks.

3.4. Malicious process behaviors

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

3.5. Online social networks and recommender systems

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

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

4. Highlights of the Year

4.1. Highlights of the Year

Anne Marie Kermarrec has been named an ACM Fellow “for her contributions to large-scale distributed computing.”

George Giakkoupis was the General Chair of the 35th ACM SIGACT-SIGOPS Symposium on Principles of Distributed Computing (PODC 2016).

Michel Raynal renewed his appointment as an Adjunct Professor at the University of Hong Kong.

4.1.1. Awards

BEST PAPERS AWARDS :

[30] **IEEE IC2E'16**. S. DELBRUEL, D. FREY, F. TAÏANI.

[27] **The International Conference on Networked Systems NETYS**. N. CHILUKA, A.-M. KERMARREC, J. OLIVARES.

5. New Software and Platforms

5.1. Asapknn (MediEgo)

KEYWORDS: Widget web - Social network - Recommendation

FUNCTIONAL DESCRIPTION

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

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

5.2. Brow2Brow

Browser-to-browser serverless toolboxes
FUNCTIONAL DESCRIPTION

Brow2Brow is an “Action de Development Technologique”, i.e. a collaborative development project that aims at providing a middleware and software library for browser-to-browser applications. Brow2Brow involves the ASAP team as well as the DICE Team from Inria Grenoble (Antenne de Lyon). The project seeks to provide an alternative to the current model followed by Web2.0 applications by exploiting the recently introduced WebRTC standard. Existing Web 2.0 applications collect data on browsers and send it to servers that store and process it. The goal of Brow2Brow is to provide an alternative approach where browsers can themselves proceed to collaborative data processing. This will make it possible avoid data concentration at a single server. The project has resulted so far in the development of WebGC, a library for gossip-based applications on browsers.

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

5.3. Dashboard

MediEgo Dashboard: A personalized news dashboard
KEYWORDS: Recommender system - Personalized stream of news - Dashboard
FUNCTIONAL DESCRIPTION

This work has led to the development of MEDIEGO Dashboard, a personalized news recommendation system. In MEDIEGO Dashboard, users benefit from a personalized stream of news matching their interests. Additionally, users can use explicit subscriptions as well as post content and navigate through tags. MEDIEGO Dashboard is available through a web interface and a mobile-based Android application. To provide personalization, MEDIEGO Dashboard exploits the users’ opinions regarding their received news to identify users with similar interests. MEDIEGO Dashboard is centralized and it allows us to test and evaluate different recommendation schemes. In collaboration with EIT/ICT Lab, an experiment has been conducted with a set of users at Trento (Italie). This experiment allowed us to collect traces and to perform a user survey to assess and improve our solution. This solution will soon be interconnected to P2P-AllYours.

- Participants: Anne-Marie Kermarrec, Antoine Boutet, Yuri Barssi and Jean Francois Verdonck
- Contact: Anne-Marie Kermarrec
- URL: http://www.mediego.com

5.4. GossipLib

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

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

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

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

5.5. HEAP

Heterogeneity-aware gossip protocol
FUNCTIONAL DESCRIPTION

A video streaming platform based on HEAP. The platform is particularly suited for environment characterized by heterogeneous bandwidth capabilities such as those comprising ADSL edge nodes. HEAP is, in fact, able to dynamically leverage the most capable nodes and increase their contribution to the protocol, while decreasing by the same proportion that of less capable nodes. During the last few months, we have integrated HEAP with the ability to dynamically measure the available bandwidth of nodes, thereby making it independent of the input of the user.

- Participants: Davide Frey, Arnaud Jegou, Anne-Marie Kermarrec, Vivien Quema, Maxime Monod and Rachid Guerraoui
- Contact: Davide Frey

5.6. HyRec

Hybrid Recommender System
FUNCTIONAL DESCRIPTION

The motivation of this work is to explore solutions that could in some sense democratize personalization by making it accessible to any content provider company without generating huge investments. HyRec implements a user-based collaborative filtering scheme and offloads CPU-intensive recommendation tasks to front-end client browsers, while retaining storage and orchestration tasks within back-end servers. HyRec seeks to provide the scalability of p2p approaches without forcing content providers to give up the control of the system.

- Participants: Antoine Boutet, Davide Frey, Anne Marie Kermarrec, Arnaud Jegou and Rachid Guerraoui
- Contact: Davide Frey

5.7. WebGC

Web-based Gossip Communication
FUNCTIONAL DESCRIPTION

WebGC is a library for gossip-based communication between web-browsers. It has been developed in collaboration with Mathieu Simonin in the context of the Brow2Brow ADT project. WebGC builds on the recent WebRTC standard as well as on PeerJS, an open-source project that provides primitives for data transfer on top of WebRTC.

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

5.8. WhatsUp

KEYWORD: Recommender system
FUNCTIONAL DESCRIPTION

WhatsUp is a distributed recommendation system aimed to distribute instant news in a large scale dynamic system. WhatsUp has two parts, an embedded application server in order to exchange with others peers in the system and a fully dynamic web interface for displaying news and collecting opinions about what the user reads. Underlying this web-based application lies Beep, a biased epidemic dissemination protocol that delivers news to interested users in a fast manner while limiting spam. Beep is parametrized on the fly to manage the orientation and the amplification of news dissemination. Every user forwards the news of interest to a randomly selected set of users with a preference towards those that have similar interests (orientation). The notion of interest does not rely on any explicit social network or subscription scheme, but rather on an implicit and dynamic overlay capturing the commonalities between users with respect to they are interested in. The size of the set of users to which a news is forwarded depends on the interest of the news (amplification). A centralized version of WhatsUp is already up and running and the decentralized one is still in beta version.

- Participants: Davide Frey, Ribeiro Heverson, Antoine Boutet, Anne Marie Kermarrec, Arnaud Jegou, Rachid Guerraoui and Jean Francois Verdonck
- Contact: Davide Frey

5.9. YALPS

KEYWORDS: Traffic-shaping - Nat traversal - Experimentation - Peer-to-peer - Simulator - Deployment
FUNCTIONAL DESCRIPTION

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

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

5.10. P2P-AllYours

Peer-to-Peer AllYours

FUNCTIONAL DESCRIPTION

P2P-AllYours is customization of WhatsUp developed in the context of the EIT/ICT-Labs AllYours project. In addition to WhatsUp (the distributed recommender engine), P2P-AllYours comprises the following features:

- A new web interface, which users can access through a local web-server integrated in P2P-AllYours.
- A set of automatic nodes (BOTs) that can extract news items from RSS feeds and insert them into the recommender system.
- A content-bootstrap that solves the issues related to bootstrapping the recommender system when a user connects for the first time.
- An experiment management server that allows users to register for the application in the context of the testing program.
- Participants: Davide Frey, Heverson Borba Ribeiro, Raziel Carvajal Gomez, Arnaud Jegou and Anne-Marie Kermarrec
- Contact: Davide Frey

6. New Results

6.1. Theory of Distributed Systems

6.1.1. t-Resilient Immediate Snapshot is Impossible

Participant: Michel Raynal.

Immediate snapshot is the basic communication object on which relies the read/write distributed computing model made up of n crash-prone asynchronous processes, called iterated distributed model. Each iteration step (usually called a round) uses a new immediate snapshot object, which allows the processes to communicate and cooperate. More precisely, the x-th immediate snapshot object can be used by a process only when it executes the x-th round. An immediate snapshot object can be implemented by an $(n{-}1)$-resilient algorithm, i.e. an algorithm that tolerates up to $(n{-}1)$ process crashes (also called wait-free algorithm). Considering a t-crash system model (i.e. a model in which up to t processes are allowed to crash), this work [46] is on the construction of an extension of immediate snapshot objects to t-resiliency. In the t-crash system model, at each round each process may be ensured to get values from at least $n{-}t$ processes, and t-immediate snapshot has the properties of classical immediate snapshot (1-immediate snapshot) but ensures that each process will get values form at least $n{-}t$ processes. Its main result is the following. While there is a (deterministic) t-resilient read/write-based algorithm implementing t-immediate snapshot in a t-crash system when $t = n{-}1$, there is no t-resilient algorithm in a t-crash model when $t \in [1..(n{-}2)]$. This means that the notion of t-resilience is inoperative when one has to implement immediate snapshot for these values of t: the model assumption "at most $t < n{-}1$ processes may crash" does not provide us with additional computational power allowing for the design of genuine t-resilient algorithms (genuine meaning that such a t-resilient algorithm would work in the t-crash model, but not in the $(t + 1)$-crash model). To show these results, the paper relies on well-known distributed computing agreement problems such as consensus and k-set agreement.

This work was done in collaboration with Carole Delporte, Hugues Fauconnier, and Sergio Rajsbaum, and appeared at SIROCCO 2016.

6.1.2. Two-Bit Messages are Sufficient to Implement Atomic Read/Write Registers in Crash-Prone Systems

Participant: Michel Raynal.

Atomic registers are certainly the most basic objects of computing science. Their implementation on top of an n-process asynchronous message-passing system has received a lot of attention. It has been shown that $t < n/2$ (where t is the maximal number of processes that may crash) is a necessary and sufficient requirement to build an atomic register on top of a crash-prone asynchronous message-passing system. Considering such a context, this work [49] presents an algorithm which implements a single-writer multi-reader atomic register with four message types only, and where no message needs to carry control information in addition to its type. Hence, two bits are sufficient to capture all the control information carried by all the implementation messages. Moreover, the messages of two types need to carry a data value while the messages of the two other types carry no value at all. As far as we know, this algorithm is the first with such an optimality property on the size of control information carried by messages. It is also particularly efficient from a time complexity point of view.

This work was done in collaboration with Achour Mostefaoui, and appeared at PODC 2016.

6.2. Network and Graph Algorithms

6.2.1. Vertex Coloring with Communication and Local Memory Constraints in Synchronous Broadcast Networks

Participants: Hicham Lakhlef, Michel Raynal, Francois Taiani.

This work [41] considers the broadcast/receive communication model in which message collisions and message conflicts can occur because processes share frequency bands. (A collision occurs when, during the same round, messages are sent to the same process by too many neighbors. A conflict occurs when a process and one of its neighbors broadcast during the same round.) More precisely, this work considers the case where, during a round, a process may either broadcast a message to its neighbors or receive a message from at most m of them. This captures communication-related constraints or a local memory constraint stating that, whatever the number of neighbors of a process, its local memory allows it to receive and store at most m messages during each round. This work defines first the corresponding generic vertex multi-coloring problem (a vertex can have several colors). It focuses then on tree networks, for which it presents a lower bound on the number of colors K that are necessary (namely, $K = \lceil \frac{\Delta}{m} \rceil + 1$, where Δ is the maximal degree of the communication graph), and an associated coloring algorithm, which is optimal with respect to K.

6.2.2. Optimal Collision/Conflict-Free Distance-2 Coloring in Wireless Synchronous Broadcast/Receive Tree Networks

Participants: Davide Frey, Hicham Lakhlef, Michel Raynal.

We studied the problem of decentralized distance-2 coloring in message-passing systems where communication is (a) synchronous and (b) based on the “broadcast/receive” pair of communication operations. “Synchronous” means that time is discrete and appears as a sequence of time slots (or rounds) such that each message is received in the very same round in which it is sent. “Broadcast/receive” means that during a round a process can either broadcast a message to its neighbors or receive a message from one of them. In such a communication model, no two neighbors of the same process, nor a process and any of its neighbors, must be allowed to broadcast during the same time slot (thereby preventing message collisions in the first case, and message conflicts in the second case). From a graph theory point of view, the allocation of slots to processes is know as the distance-2 coloring problem: a color must be associated with each process (defining the time slots in which it will be allowed to broadcast) in such a way that any two processes at distance at most 2 obtain different colors, while the total number of colors is “as small as possible”. In this context, we proposed a parallel message-passing distance-2 coloring algorithm suited to trees, whose roots are dynamically defined. This algorithm, which is itself collision-free and conflict-free, uses $\Delta + 1$ colors where Δ is the maximal degree of the graph (hence the algorithm is color-optimal). It does not require all processes to have different initial identities, and its time complexity is $O(d\Delta)$, where d is the depth of the tree. As far as we know, this is the first distributed distance-2 coloring algorithm designed for the broadcast/receive round-based communication model, which owns all the previous properties. We published these results in [39].

6.2.3. Efficient Plurality Consensus, or: The Benefits of Cleaning Up from Time to Time

Participant: George Giakkoupis.

Plurality consensus considers a network of n nodes, each having one of k opinions. Nodes execute a (randomized) distributed protocol with the goal that all nodes adopt the *plurality* (the opinion initially supported by the most nodes). Communication is realized via the random phone call model. A major open question has been whether there is a protocol for the complete graph that converges (w.h.p.) in polylogarithmic time and uses only polylogarithmic memory per node (local memory). We answered this question affirmatively.

In [22], we propose two protocols that need only mild assumptions on the bias in favor of the plurality. As an example of our results, consider the complete graph and an arbitrarily small constant multiplicative bias in favor of the plurality. Our first protocol achieves plurality consensus in $O(\log k \cdot \log\log n)$ rounds using $\log k + O(\log\log k)$ bits of local memory. Our second protocol achieves plurality consensus in $O(\log n \cdot \log\log n)$ rounds using only $\log k + 4$ bits of local memory. This disproves a conjecture by Becchetti et al. (SODA'15) implying that any protocol with local memory $\log k + O(1)$ has worst-case runtime $\Omega(k)$. We provide similar bounds for much weaker bias assumptions. At the heart of our protocols lies an *undecided state*, an idea introduced by Angluin et al. (Distributed Computing'08).

This work was done in collaboration with Petra Berenbrink (SFU), Tom Friedetzky (Durham University), and Peter Kling (SFU).

6.2.4. Bounds on the Voter Model in Dynamic Networks

Participants: George Giakkoupis, Anne-Marie Kermarrec.

In the *voter model*, each node of a graph has an opinion, and in every round each node chooses independently a random neighbour and adopts its opinion. We are interested in the *consensus time*, which is the first point in time where all nodes have the same opinion. In [23], we consider dynamic graphs in which the edges are rewired in every round (by an adversary) giving rise to the graph sequence $G_1, G_2, \cdots$, where we assume that G_i has conductance at least ϕ_i. We assume that the degrees of nodes don't change over time as one can show that the consensus time can become super-exponential otherwise. In the case of a sequence of d-regular graphs, we obtain asymptotically tight results. Even for some static graphs, such as the cycle, our results improve the state of the art. Here we show that the expected number of rounds until all nodes have the same opinion is bounded by $O(m/(\delta \cdot \phi))$, for any graph with m edges, conductance ϕ, and degrees at least δ. In addition, we consider a *biased* dynamic voter model, where each opinion i is associated with a probability P_i, and when a node chooses a neighbour with that opinion, it adopts opinion i with probability P_i (otherwise the node keeps its current opinion). We show for any regular dynamic graph, that if there is an $\epsilon > 0$ difference between the highest and second highest opinion probabilities, and at least $\Omega(\log n)$ nodes have initially the opinion with the highest probability, then all nodes adopt w.h.p. that opinion. We obtain a bound on the convergence time, which becomes $O(\log n/\phi)$ for static graphs.

This work was done in collaboration with Petra Berenbrink (SFU), and Frederik Mallmann-Trenn (SFU).

6.2.5. How Asynchrony Affects Rumor Spreading Time

Participant: George Giakkoupis.

In standard randomized (push-pull) rumor spreading, nodes communicate in synchronized rounds. In each round every node contacts a random neighbor in order to exchange the rumor (i.e., either push the rumor to its neighbor or pull it from the neighbor). A natural asynchronous variant of this algorithm is one where each node has an independent Poisson clock with rate 1, and every node contacts a random neighbor whenever its clock ticks. This asynchronous variant is arguably a more realistic model in various settings, including message broadcasting in communication networks, and information dissemination in social networks.

In [35] we study how asynchrony affects the rumor spreading time, that is, the time before a rumor originated at a single node spreads to all nodes in the graph. Our first result states that the asynchronous push-pull rumor spreading time is asymptotically bounded by the standard synchronous time. Precisely, we show that for any graph G on n-nodes, where the synchronous push-pull protocol informs all nodes within $T(G)$ rounds with high probability, the asynchronous protocol needs at most time $O(T(G) + \log n)$ to inform all nodes with high probability. On the other hand, we show that the expected synchronous push-pull rumor spreading time is bounded by $O(\sqrt{n})$ times the expected asynchronous time.

These results improve upon the bounds for both directions shown recently by Acan et al. (PODC 2015). An interesting implication of our first result is that in regular graphs, the weaker push-only variant of synchronous rumor spreading has the same asymptotic performance as the synchronous push-pull algorithm.

This work was done in collaboration with Yasamin Nazari and Philipp Woelfel from the University of Calgary.

6.2.6. Amplifiers and Suppressors of Selection for the Moran Process on Undirected Graphs

Participant: George Giakkoupis.

In [47] we consider the classic Moran process modeling the spread of genetic mutations, as extended to structured populations by Lieberman et al. (Nature, 2005). In this process, individuals are the vertices of a connected graph G. Initially, there is a single mutant vertex, chosen uniformly at random. In each step, a random vertex is selected for reproduction with a probability proportional to its fitness: mutants have fitness $r > 1$, while non-mutants have fitness 1. The vertex chosen to reproduce places a copy of itself to a uniformly random neighbor in G, replacing the individual that was there. The process ends when the mutation either reaches fixation (i.e., all vertices are mutants), or gets extinct. The principal quantity of interest is the probability with which each of the two outcomes occurs.

A problem that has received significant attention recently concerns the existence of families of graphs, called strong amplifiers of selection, for which the fixation probability tends to 1 as the order n of the graph increases, and the existence of strong suppressors of selection, for which this probability tends to 0. For the case of directed graphs, it is known that both strong amplifiers and suppressors exist. For the case of undirected graphs, however, the problem has remained open, and the general belief has been that neither strong amplifiers nor suppressors exist. In this work we disprove this belief, by providing the first examples of such graphs. The strong amplifier we present has fixation probability $1 - \widetilde{O}(n^{-1/3})$, and the strong suppressor has fixation probability $\widetilde{O}(n^{-1/4})$. Both graph constructions are surprisingly simple. We also prove a general upper bound of $1 - \widetilde{\Omega}(n^{-1/3})$ on the fixation probability of any undirected graph. Hence, our strong amplifier is existentially optimal.

6.3. Scalable Systems

6.3.1. Cache locality is not enough: High-Performance Nearest Neighbor Search with Product Quantization Fast Scan

Participants: Fabien Andre, Anne-Marie Kermarrec.

Nearest Neighbor (NN) search in high dimension is an important feature in many applications (e.g., image retrieval, multimedia databases). Product Quantization (PQ) is a widely used solution which offers high performance, i.e., low response time while preserving a high accuracy. PQ represents high-dimensional vectors (e.g., image descriptors) by compact codes. Hence, very large databases can be stored in memory, allowing NN queries without resorting to slow I/O operations. PQ computes distances to neighbors using cache-resident lookup tables, thus its performance remains limited by (i) the many cache accesses that the algorithm requires, and (ii) its inability to leverage SIMD instructions available on modern CPUs. In this paper, we advocate that cache locality is not sufficient for efficiency. To address these limitations, in [19] we design a novel algorithm, PQ Fast Scan, that transforms the cache-resident lookup tables into small tables, sized to fit SIMD registers. This transformation allows (i) in-register lookups in place of cache accesses and (ii) an efficient SIMD implementation. PQ Fast Scan has the exact same accuracy as PQ, while having 4 to 6 times lower response time (e.g., for 25 million vectors, scan time is reduced from 74ms to 13ms).

6.3.2. Toward an Holistic Approach of Systems-of-Systems

Participants: Simon Bouget, David Bromberg, Francois Taiani.

Large scale distributed systems have become ubiquitous, from on-line social networks to the Internet-of-Things. To meet rising expectations (scalability, robustness, flexibility,...) these systems increasingly espouse complex distributed architectures, that are hard to design, deploy and maintain. To grasp this complexity, developers should be allowed to assemble large distributed systems from smaller parts using a seamless, high-level programming paradigm. We present in [24] such an assembly-based programming framework, enabling developers to easily define and realize complex distributed topologies as a construction of simpler blocks (e.g. rings, grids). It does so by harnessing the power of self-organizing overlays, that is made accessible to developers through a high-level Domain Specific Language and self-stabilizing run-time. Our evaluation further shows that our approach is generic, expressive, low-overhead and robust.

6.3.3. Speed for the Elite, Consistency for the Masses: Differentiating Eventual Consistency in Large-Scale Distributed Systems

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

Eventual consistency is a consistency model that emphasizes liveness over safety; it is often used for its ability to scale as distributed systems grow larger. Eventual consistency tends to be uniformly applied to an entire system, but we argue that there is a growing demand for differentiated eventual consistency requirements.

We address this demand with UPS [34], a novel consistency mechanism that offers differentiated eventual consistency and delivery speed by working in pair with a two-phase epidemic broadcast protocol. We propose a closed-form analysis of our approach's delivery speed, and we evaluate our complete mechanism experimentally on a simulated network of one million nodes. To measure the consistency trade-off, we formally define a novel and scalable consistency metric that operates at runtime. In our simulations, UPS divides by more than 4 the inconsistencies experienced by a majority of the nodes, while reducing the average latency incurred by a small fraction of the nodes from 6 rounds down to 3 rounds.

This work was done in collaboration with Achour Mostefaoui and Matthieu Perrin from the LINA laboratory in Nantes.

6.3.4. Bringing Secure Bitcoin Transactions to your Smartphone

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

To preserve the Bitcoin ledger's integrity, a node that joins the system must download a full copy of the entire Bitcoin blockchain if it wants to verify newly created blocks. At the time of writing, the blockchain weights 79 GiB and takes hours of processing on high-end machines. Owners of low-resource devices (known as thin nodes), such as smartphones, avoid that cost by either opting for minimum verification or by depending on full nodes, which weakens their security model.

In this work [33], we propose to harden the security model of thin nodes by enabling them to verify blocks in an adaptive manner, with regards to the level of targeted confidence, with low storage requirements and a short bootstrap time. Our approach exploits sharing within a distributed hash table (DHT) to distribute the storage load, and a few additional hashes to prevent attacks on this new system.

This work was done in collaboration with Marc X. Makkes and Spyros Voulgaris from Vrije Universiteit Amsterdam (The Netherlands).

6.3.5. Multithreading Approach to Process Real-Time Updates in KNN Algorithms

Participants: Anne-Marie Kermarrec, Nupur Mittal, Javier Olivares.

K-Nearest Neighbors algorithm is the core of a considerable amount of online services and applications, like recommendation engines, content-classifiers, information retrieval systems, etc. The users of these services change their preferences and evolve with time, aggravating the computational challenges of KNN more with the ever evolving data to process. In this work [48], we present *UpKNN*: an efficient thread-based approach to take the updates of users preferences into account while it computes the KNN efficiently, keeping a check on the wall-time.

By using an efficient thread-based approach, *UpKNN* processes millions of updates online, on a single commodity PC. Our experiments confirm the scalability of *UpKNN*, both in terms of the number of updates processed and the threads used. *UpKNN* achieves speedups ranging from 13.64X to 49.5X in the processing of millions of updates, with respect to the performance of a non-partitioned baseline. These results are a direct consequence of reducing the number of disk operations, roughly speaking, only 1% disk operations are performed as compared to the baseline.

6.3.6. The Out-of-Core KNN Awakens: The Light Side of Computation Force on Large Datasets

Participants: Anne-Marie Kermarrec, Javier Olivares.

K-Nearest Neighbors (KNN) is a crucial tool for many applications, e.g. recommender systems, image classification and web-related applications. However, KNN is a resource greedy operation particularly for large datasets. We focus on the challenge of KNN computation over large datasets on a single commodity PC with limited memory. We propose a novel approach [27] to compute KNN on large datasets by leveraging both disk and main memory efficiently. The main rationale of our approach is to minimize random accesses to disk, maximize sequential accesses to data and efficient usage of only the available memory.

We evaluate our approach on large datasets, in terms of performance and memory consumption. The evaluation shows that our approach requires only 7% of the time needed by an in-memory baseline to compute a KNN graph.

6.3.7. Partial Replication Policies for Dynamic Distributed Transactional Memory in Edge Clouds

Participant: Francois Taiani.

Distributed Transactional Memory (DTM) can play a fundamental role in the coordination of participants in edge clouds as a support for mobile distributed applications. DTM emerges as a concurrency mechanism aimed at simplifying distributed programming by allowing groups of operations to execute atomically, mirroring the well-known transaction model of relational databases. In spite of recent studies showing that partial replication approaches can present gains in the scalability of DTMs by reducing the amount of data stored at each node, most DTM solutions follow a full replication scheme. The few partial replicated DTM frameworks either follow a random or round-robin algorithm for distributing data onto partial replication groups. In order to overcome the poor performance of these schemes, this work [36] investigates policies to extend the DTM to efficiently and dynamically map resources on partial replication groups. The goal is to understand if a dynamic service that constantly evaluates the data mapped into partial replicated groups can contribute to improve DTM based systems performance.

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

6.3.8. Being Prepared in a Sparse World: The Case of KNN Graph Construction

Participants: Anne-Marie Kermarrec, Nupur Mittal, Francois Taiani.

Work [25] presents KIFF, a generic, fast and scalable KNN graph construction algorithm. KIFF directly exploits the bipartite nature of most datasets to which KNN algorithms are applied. This simple but powerful strategy drastically limits the computational cost required to rapidly converge to an accurate KNN solution, especially for sparse datasets. Our evaluation on a representative range of datasets show that KIFF provides, on average, a speed-up factor of 14 against recent state-of-the art solutions while improving the quality of the KNN approximation by 18

This work was done in collaboration with Antoine Boutet from CNRS, Laboratoire Hubert Curien, Saint-Etienne, France.

6.3.9. *Exploring the Use of Tags for Georeplicated Content Placement*

Participants: Stephane Delbruel, Davide Frey, Francois Taiani.

A large portion of today's Internet traffic originates from streaming and video services. Such services rely on a combination of distributed datacenters, powerful content delivery networks (CDN), and multi-level caching . In spite of this infrastructure, storing, indexing, and serving these videos remains a daily engineering challenge that requires increasing efforts on the part of providers and ISPs. In this work [30], we explore how the tags attached to videos by users could help improve this infrastructure, and lead to better performance on a global scale. Our analysis shows that tags can be interpreted as markers of a video's geographic diffusion, with some tags strongly linked to well identified geographic areas. Based on our findings, we demonstrate the potential of tags to help predict distribution of a video's views, and present results suggesting that tags can help place videos in globally distributed datacenters. We show in particular that even a simplistic approach based on tags can help predict a minimum of 65.9% of a video's views for a majority of videos, and that a simple tag-based placement strategy is able to improve the hit rate of a distributed on-line video service by up to 6.8% globally over a naive random allocation.

6.3.10. *Mignon: A Fast Decentralized Content Consumption Estimation in Large-Scale Distributed Systems*

Participants: Stephane Delbruel, Davide Frey, Francois Taiani.

Although many fully decentralized content distribution systems have been proposed, they often lack key capabilities that make them difficult to deploy and use in practice. In this work [31], we look at the particular problem of content consumption prediction, a crucial mechanism in many such systems. We propose a novel, fully decentralized protocol that uses the tags attached by users to on-line content, and exploits the properties of self-organizing kNN overlays to rapidly estimate the potential of a particular content without explicit aggregation.

6.4. Privacy in User Centric Applications

6.4.1. *Hybrid Recommendations with Dynamic Similarity Measure*

Participants: Anne-Marie Kermarrec, Nupur Mittal.

This project aims to combine the classical methods of content based and collaborative filtering recommendations, in addition to dynamic similarity computations. The objective is to exploit the varied item-data available from the world wide web, to overcome trivial problems like that of cold-start. In this work, we have designed a new similarity metric inspired from the existing DICE similarity that takes into account changing item/user behavior to compute updated similarity values for the purpose of recommendations. The work leverages the idea of content based recommendations as a first step to create vivid user and item profiles that are iteratively updated.

This work was done in collaboration with Rachid Guerraoui (EPFL, Switzerland), Rhicheek Patra (EPFL, Switzerland).

6.4.2. *Lightweight Privacy-Preserving Averaging for the Internet of Things*

Participants: Davide Frey, George Giakkoupis, Julien Lepiller.

The number of connected devices is growing continuously, and so is their presence into our everyday lives. From GPS-enabled fitness trackers, to smart fridges that tell us what we need to buy at the grocery store, connected devices—things—have the potential to collect and make available significant amounts of information. On the one hand, this information may provide useful services to users, and constitute a statistical gold mine. On the other, its availability poses serious privacy threats for users. In this work, we designed two new protocols that make it possible to aggregate personal information collected by smart devices in the form of an average, while preventing attackers from learning the details of the non-aggregated data. The first protocol exploits randomness and decomposition into shares as techniques to obfuscate the value associated with each node and lightweight encryption techniques to withstand eavesdropping attacks. The second exploits only randomness and encryption. We carried out a preliminary evaluation and published the results related to the first protocol in [18].

This work was done in collaboration with Tristan Allard from the DRUID Team at IRISA, Rennes.

6.4.3. Collaborative Filtering Under a Sybil Attack: Similarity Metrics do Matter!

Participants: Davide Frey, Anne-Marie Kermarrec, Antoine Rault, Florestan de Moor.

Whether we are shopping for an interesting book or selecting a movie to watch, the chances are that a recommendation system will help us decide what we want. Recommendation systems collect information about our own preferences, compare them to those of other users, and provide us with suggestions on a variety of topics. But is the information gathered by a recommendation system safe from potential attackers, be them other users, or companies that access the recommendation system? And above all, can service providers protect this information while still providing effective recommendations? In this work, we analyze the effect of Sybil attacks on collaborative-filtering recommendation systems, and discuss the impact of different similarity metrics in the trade-off between recommendation quality and privacy. Our results, on a state-of-the-art recommendation framework and on real datasets show that existing similarity metrics exhibit a wide range of behaviors in the presence of Sybil attacks. Yet, they are all subject to the same trade off: Sybil resilience for recommendation quality. We therefore propose and evaluate a novel similarity metric that combines the best of both worlds: a low RMSE score with a prediction accuracy for Sybil users of only a few percent. A preliminary version of this work was published at EuroSec 2015 [57]. This year, we significantly extended the work during the summer internship of Florestan De Moor. Specifically, we considered new attacks that specifically target our novel similarity metric and showed that regardless of the attack configuration, our metric can preserve the privacy of users without hampering recommendation quality. A new paper with these new results was submitted to PETS 2017.

6.4.4. Privacy-Preserving Distributed Collaborative Filtering

Participants: Davide Frey, Anne-Marie Kermarrec.

In this work, we propose a new mechanism to preserve privacy while leveraging user profiles in distributed recommender systems. Our mechanism relies on (*i*) an original obfuscation scheme to hide the exact profiles of users without significantly decreasing their utility, as well as on (*ii*) a randomized dissemination protocol ensuring differential privacy during the dissemination process.

We compare our mechanism with a non-private as well as with a fully private alternative. We consider a real dataset from a user survey and report on simulations as well as planetlab experiments. We dissect our results in terms of accuracy and privacy trade-offs, bandwith consumption, as well as resilience to a censorship attack. In short, our extensive evaluation shows that our twofold mechanism provides a good trade-off between privacy and accuracy, with little overhead and high resilience.

This work was done with Antoine Boutet and Arnaud Jegou when they were part of the team, and in collaboration with Rachid Guerraoui from EPFL. But the complete results were published this year in [15].

7. Bilateral Contracts and Grants with Industry

7.1. Contract with Technicolor

Participants: Fabien Andre, Anne-Marie Kermarrec.

We had a contract with Technicolor for collaboration on large-scale infrastructure for recommendation systems. In this context, Anne-Marie Kermarrec was the PhD advisor of Fabien Andre until Nov 2016. In his Phd, Fabien Andre worked on efficient algorithms for heterogeneous data on large-scale platforms.

8. Partnerships and Cooperations

8.1. National Initiatives

8.1.1. ANR project SocioPlug

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

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

8.1.2. DeSceNt CominLabs

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

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

8.1.3. ANR Blanc project Displexity

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

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

8.1.4. ANR project PAMELA

Participants: Davide Frey, George Giakkoupis, Francois Taiani.

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

8.1.5. ANR project OBrowser

Participants: David Bromberg, Davide Frey, Francois Taiani.

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

8.1.6. ANR project DESCARTES

Participants: George Giakkoupis, Michel Raynal, Francois Taiani.

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

8.1.7. ANR-ERC Tremplin project NDFUSION

Participant: George Giakkoupis.

NDFUSION is an 18-month ANR project awarded to the PI to support his preparation for his upcoming ERC grant application. The idea of intervening in a network diffusion process to enhance or retard its spread has been studied in various contexts, e.g., to increase the spread or speed of diffusion by choosing an appropriate set of seed nodes (a standard goal in viral marketing by word-of-mouth), or achieve the opposite effect either by choosing a small set of nodes to remove (a goal in immunization against diseases), or by seeding a competing diffusion (e.g., to limit the spread of misinformation in a social network). The aim of this project is to consolidate existing work under a single, comprehensive framework, and using this framework to develop new, efficient algorithms for optimizing (maximizing or minimizing) the spread of diffusion processes. Novel aspects of the project involve issues of scalability, multiple concurrent diffusions, and the use of multistage online strategies to optimize diffusions. Results from this project are likely to be relevant to many different disciplines, from network optimization in computing to disease containment in medicine.

8.2. International Initiatives

8.2.1. Inria International Labs

Anne-Marie Kermarrec has been scientific collaborator at EPFL, Lausanne, since Feb 2014.

Anne-Marie Kermarrec has been the scientific coordinator of the EPFL/Inria International Lab since Feb 2015.

8.3. International Research Visitors

8.3.1. Visits of International Scientists

Antonio Carzaniga, Università della Svizzera italiana (USI), Apr 29.

Evangelos Bampas, Aix-Marseille Université, Mar 23.

Fábio Moreira Costa, Institute of Informatics, Federal University of Goiás, Goiânia-GO, Brazil, Sep 12–16.

Ricardo Couto Antunes da Rocha, Department of Computer Science, Federal University of Goiás, Catalão-GO, Brazil, Sep 12–16.

Rachid Guerraoui, EPFL, Switzerland, invited Professor at Univ. Rennes I / ISTIC, since September 2016.

Arvid Bosk, KTH Royal Institute of Technology, guest PhD Student, from Dec 2016.

8.3.1.1. Internships

Florestan De Moor, Collaborative Filtering Under a Sybil Attack, Univ. Rennes I, May to Jul 2016, supervised by Davide Frey.

Julien Lepiller, Private Decentralized Aggregation, Inria, Feb to Jun 2016, supervised by Davide Frey and Francois Taiani.

8.3.2. Visits to International Teams

8.3.2.1. Research Stays Abroad

David Bromberg visited UFG, Goiania, Brazil, from Jun to Jul 2016 (CONFAP CNRS project)

8.3.2.2. Internships

Pierre-Louis Roman did an internship at Vrije Universiteit Amsterdam, The Netherlands, from Jun to Aug 2016 under the supervision of Spyros Voulgaris, with a grant from Université Bretagne Loire.

Simon Bouget did an internship at Centre for Complex Systems and Big Data at University of Neuchatel, Switzerland, from May to Aug 2016, under the supervision of Etienne Rivière, with a grant from the “Outgoing Mobility for Doctoral Students” program of Rennes Métropole.

Stéphane Delbruel did an internship at Universita della Svizzera Italiana, Switzerland, from Jun to Jul 2016, under the supervision of Antonio Carzaniga.

9. Dissemination

9.1. Promoting Scientific Activities

9.1.1. Scientific Events Organisation

9.1.1.1. General Chair, Scientific Chair

George Giakkoupis was the General Chair of the 35th ACM SIGACT-SIGOPS Symposium on Principles of Distributed Computing (PODC 2016), Chicago, IL, USA, Jul 2016.

9.1.2. Scientific Events Selection

9.1.2.1. Member of Conference Program Committees

Pierre-Louis Roman served in the program committee of the 6th IEEE International Symposium on Cloud and Service Computing (SC2), Nadi, Fiji, Dec 2016.

Francois Taiani served in the program committee of the:

17th ACM/IFIP/USENIX International Conference on Middleware (Middleware), Trento, Italy, Dec 2016.

15th Workshop on Adaptive and Reflective Middleware (ARM’16), Trento, Italy, Dec 2016.

4th workshop on Distributed Cloud Computing (DCC), Chicago, USA, Jul 2016

International Workshop on Assurance in Distributed Systems and Networks (ADSN 2016), Nara, Japan, Jun 2016.

IEEE International Conference on Cloud Engineering 2016 (IC2E 2016), Berlin, Germany, Apr 2016.

David Bromberg served in the program committee of the:

17th ACM/IFIP/USENIX International Conference on Middleware (Middleware), Trento, Italy, Dec 2016.

IEEE International Conference on Cloud Engineering 2016 (IC2E 2016), Berlin, Germany, Apr 2016.

Davide Frey served in the program committee of the

11th ACM International Conference on Distributed and Event-Based Systems (DEBS 2017), Barcelona, Spain, Jun 2017.

17th ACM/IFIP/USENIX International Conference on Middleware (Middleware), Trento, Italy, Dec 2016.

George Giakkoupis served in the program committee of the:

17th International Conference on Distributed Computing and Networking (ICDCN 2017), Singapore, Jan 2016.

3rd ACM-W Europe Celebration of Women in Computing (womENcourage 2016), Linz, Austria, Sep 2016.

31st IEEE International Parallel & Distributed Processing Symposium (IPDPS 2017), Orlando, FL, USA, May 2017.

9.1.3. Journal

9.1.3.1. Member of Editorial Boards

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

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

9.1.4. Invited Talks

Michel Raynal, A Look at Basics of Distributed Computing, Keynote Talk in the 36th International Conference on Distributed Computing Systems (ICDCS 2016), Nara, Japan, June 2016.

Pierre-Louis Roman, Bringing secure Bitcoin transactions to your smartphone, Technicolor, Rennes, Nov 24 2016.

Francois Taiani, Being prepared in a sparse world: The case of KNN graph construction, Invited Seminar of the Département Informatique et télécommunications (DIT) of ENS Rennes, Sep 2016.

Francois Taiani, Being prepared in a sparse world: The case of KNN graph construction, Invited Seminar of the Inria Project Team TAO, Inria Saclay, May 2016.

David Bromberg, Towards microservices architecture to transcode videos in the large at low costs, Invited seminar of the computer science department of UFG - Instituto de Informatica, Goiâna, Brasil, Jun 2016.

George Giakkoupis, Randomized adversary models. CMO-BIRS Workshop on Complexity and Analysis of Distributed Algorithms, Oaxaca, Mexico, Dec 1 2016.

George Giakkoupis, Amplifiers and suppressors for the Moran process on undirected graphs. ART-Oberseminar, University of Hamburg, Germany, Nov 8 2016.

George Giakkoupis, Amplifiers and suppressors of evolutionary dynamics on undirected graphs. Workshop on Random Processes in Discrete Structures, University of Warwick, UK, Aug 31 2016.

George Giakkoupis, Recent advances in randomized rumor spreading. Nexus of Information and Computation Theories, Distributed Computation and Communication Theme, The Henri Poincare Institute (IHP), Paris, France, Feb 1 2016.

Davide Frey gave an Invited (Keynote) talk at the W-PSDS workshop co-located with SRDS 2016 in Budapest, Hungary, on Sep 26, 2016.

Davide Frey gave an Invited talk at the SG-2 meeting of the BDVA Summit in Valencia, Spain, on Dec 1, 2016.

9.1.5. Leadership within the Scientific Community

Anne-Marie Kermarrec has been a member of the Academia Europea since 2013.

Anne-Marie Kermarrec is a member of the scientific committee of the Société Informatique de France.

George Giakkoupis was a member of the Steering Committee of the 35th ACM SIGACT-SIGOPS Symposium on Principles of Distributed Computing (PODC 2016), Chicago, IL, USA, Jul 2016.

Francois Taiani has been a member of the Steering Committee of IFIP WG 6.1 International Conference on Distributed Applications and Interoperable Systems (DAIS) since 2013.

Francois Taiani has been a member of the Steering Committee of the ACM/IFIP/USENIX International Conference on Middleware (Middleware) since 2014.

Francois Taiani served on the hiring committee for the position of Maître de Conférence a the University of Perpignan – Via Domitia in 2016.

9.1.6. Research Administration

David Bromberg is member of the scientific committee of the Media & Networks competitiveness cluster.

David Bromberg is scientific correspondent for ICT Digital in Rennes.

Davide Frey is “correspondant scientifique Europe” at the DPEI for Inria Rennes.

Davide Frey is an associate member of the COST-GTRI of Inria.

Anne-Marie Kermarrec has been a member of the ERC panel for Consolidator Grants since 2013.

9.2. Teaching - Supervision - Juries

9.2.1. Teaching

Licence : Olivier Ruas, Initiation au Génie Logiciel (GEN), 40, L2, Univ. Rennes I (ISTIC), France

Licence : Olivier Ruas, Algorithmique Fonctionnelle (APF), 24, L1, Univ. Rennes I (ISTIC), France

Engineering School : Simon Bouget, SYS1 (Introduction aux Systèmes d’exploitation), 6 HEQTD (3x 2 heures de TP), ESIR1 (L3), Univ. Rennes I (ESIR), France

Master : Simon Bouget, TLC (Techniques Logicielles pour le Cloud computing), 18 HEQTD (7x 2 heures de TP + 2x 2 heures de TD), M2 Génie Logiciel Alternance, Univ. Rennes I (ISTIC), France

Engineering School: Francois Taiani, Synchronization and Parallel Programming, 34h, 2nd year of Engineering School (M1), ESIR / Univ. Rennes I, France.

Engineering School: Francois Taiani, Distributed Systems, 24h, 3rd year of Engineering School (M2), ESIR / Univ. Rennes I, France.

Engineering School: Francois Taiani, Parallel Algorithms for Big Data, 24h, 3rd year of Engineering School (M2), ESIR / Univ. Rennes I, France.

Engineering School: Francois Taiani, Introduction to Operating Systems, 24h, 1st year of Engineering School (L3), ESIR / Univ. Rennes I, France.

Master: Francois Taiani, Programming Technologies for the Cloud, 14h, M2, Univ. Rennes I, France.

Engineering School: David Bromberg, tools and programming languages for the Web, 48h, 2nd year of Engineering School (M1), ESIR / Univ. Rennes I, France.

Engineering School: David Bromberg, Distributed software architecture, 24h, 2nd year of Engineering School (M1), ESIR / Univ. Rennes I, France.

Engineering School: David Bromberg, Network security, 48h, 2nd year of Engineering School (M1), ESIR / Univ. Rennes I, France.

Engineering School: David Bromberg, Cloud for the Internet of Things, 24h, 3rd year of Engineering School (M2), ESIR / Univ. Rennes I, France.

Master: Davide Frey, Scalable Distributed Systems, 10 hours, M1, EIT/ICT Labs Master School, Univ. Rennes I, France.

9.2.2. Supervision

PhD: Antoine Rault, User privacy in collaborative filtering systems [13], Univ. Rennes I, Jun 23, 2016, Anne-Marie Kermarrec and Davide Frey.

PhD: Fabien Andre, Exploiting Modern Hardware for Large Scale Nearest Neighbor Search [52], [0] Nov 25 2016, Anne-Marie Kermarrec (and Nicolas Le Scouarnec–Technicolor).

PhD: Nupur Mittal, Data, Learning & Privacy in Recommendation Systems [11], Nov 25, 2016, George Giakkoupis and Anne-Marie Kermarrec.

PhD: Javier Olivares, Scaling Out-of-Core K-Nearest Neighbors Computation on Single Machines [12], Dec 19 2016, Anne-Marie Kermarrec.

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

PhD in progress : Stéphane Delbruel, Towards a Decentralized Embryomorphic Storage System, Oct 2013, Francois Taiani and Davide Frey (since Oct 2014).

PhD in progress : Resmi Ariyattu Chandrasekharannair, Towards Decentralized Federations for Plug-based Decentralized Social Networks, Dec 2013, Francois Taiani.

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

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

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

9.2.3. Juries

Francois Taiani was an internal examiner of the PhD of Jackie Bourgeois. Interactive Demand Shifting in the Context of Domestic Micro-Generation. Open University (UK) / Univ. Rennes I, Jun 30 2016.

Francois Taiani was the chair of the PhD panel of Anca Iordache. Performance-Cost Trade-Offs in Heterogeneous Clouds. Univ. Rennes I, Sep 9 2016.

Francois Taiani was a reviewer of the PhD of Milan Kabac. A Design-Driven Methodology for the Development of Large-Scale Orchestrating Applications. Univ. Bordeaux, rapporteur, Sep 26 2016.

Francois Taiani was a reviewer of the PhD of Thomas Hartmann. Enabling Model-Driven Live Analytics For Cyber-Physical Systems: The Case of Smart Grids. Univ. of Luxembourg (Luxembourg), Nov 8 2016.

[0] The thesis was not yet available at HAL at the time when this report was submitted.

Francois Taiani was a reviewer of the PhD of Lucas Perrone. Vers des protocoles de tolérance aux fautes Byzantines efficaces et robustes. Univ. Grenoble Alpes, Dec 8 2016.

Francois Taiani was the chair a of the PhD panel of Javier Olivares. Scaling out-of-core K-Nearest Neighbors Computation on Single Machines. Univ. Rennes I, Dec 19 2016.

Francois Taiani was a reviewer of the HDR of Fabien Huet. From HPC to Big Data : Models and Tools for Large Scale Middleware. Université of Nice, Sophia Antipolis, Feb 12 2016.

Francois Taiani was the chair of the HDR panel of Erwan Le Merrer. Structuring Overlays for Distributed Computing: From Edge-Devices to the Clouds. Univ. Rennes I, Nov 23 2016.

George Giakkoupis was a member of the PhD committee of Maryam Helmi. The Space Complexity of Distributed Tasks in the Shared Memory Model. University of Calgary, Jun 7 2016.

George Giakkoupis was a member of the PhD committee member of Scott Roche. Robust Local Algorithms for Communication and Stability in Distributed Networks. Northeastern University, Nov 14 2016.

George Giakkoupis was an external reviewer of the PhD dissertation of Emanuele Natale. On the Computational Power of Simple Dynamics. Sapienza University of Rome, Dec 15 2016.

10. Bibliography

Major publications by the team in recent years

[1] F. André, A.-M. Kermarrec, E. Le Merrer, N. Le Scouarnec, G. Straub, A. Van Kempen.*Archiving Cold Data in Warehouses with Clustered Network Coding*, in "EuroSys 2014", Amsterdam, Netherlands, ACM New York, NY, USA, April 2014 [*DOI :* 10.1145/2592798.2592816], https://hal.inria.fr/hal-00994660.

[2] M. Bertier, D. Frey, R. Guerraoui, A.-M. Kermarrec, V. Leroy.*The Gossple Anonymous Social Network*, in "ACM/IFIP/USENIX 11th International Middleware Conference", India Bangalore, November 2010.

[3] S. Bouget, A.-M. Kermarrec, H. Kervadec, F. Taïani.*Polystyrene: the Decentralized Data Shape That Never Dies*, in "The 2014 IEEE 34th International Conference on Distributed Computing Systems (ICDCS 2014)", Madrid, Spain, June 2014, p. 288 - 297 [*DOI :* 10.1109/ICDCS.2014.37], https://hal.inria.fr/hal-01080608.

[4] J. Cao, M. Raynal, X. Yang, W. Wu.*Design and Performance Evaluation of Efficient Consensus Protocols for Mobile Ad Hoc Networks*, in "IEEE Transactions on Computers", 2007, vol. 56, n^{o} 8, p. 1055–1070.

[5] K. Censor-Hillel, M. Ghaffari, G. Giakkoupis, B. Haeupler, F. Kuhn.*Tight Bounds on Vertex Connectivity Under Vertex Sampling*, in "26th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2015)", San Diego, CA, United States, January 2015, p. 2006-1018 [*DOI :* 10.1137/1.9781611973730.133], https://hal.inria.fr/hal-01250519.

[6] G. Giakkoupis, P. Woelfel.*Randomized Mutual Exclusion with Constant Amortized RMR Complexity on the DSM*, in "55th IEEE Symposium on Foundations of Computer Science (FOCS)", Philadelphia, PA, United States, October 2014, https://hal.inria.fr/hal-01097144.

[7] M. Jelasity, S. Voulgaris, R. Guerraoui, A.-M. Kermarrec, M. van Steen.*Gossip-Based Peer Sampling*, in "ACM Transactions on Computer Systems", August 2007, vol. 41, n^{o} 5.

[8] A.-M. KERMARREC, F. TAÏANI, J. M. TIRADO MARTIN.*Scaling Out Link Prediction with SNAPLE*, in "16th Annual ACM/IFIP/USENIX Middleware Conference", Vancouver, Canada, December 2015, 12 [*DOI :* 10.1145/2814576.2814810], https://hal.inria.fr/hal-01244663.

[9] A. MOSTÉFAOUI, H. MOUMEN, M. RAYNAL.*Signature-Free Asynchronous Binary Byzantine Consensus with t< n/3, O(n*^{2}*) Messages, and O(1) Expected Time*, in "Journal of the ACM (JACM)", December 2015, vol. 62, p. 1000-1020 [*DOI :* 10.1145/2785953], https://hal.archives-ouvertes.fr/hal-01176110.

[10] F. TAÏANI, S. LIN, G. S. BLAIR.*GOSSIPKIT: A Unified Component Framework for Gossip*, in "IEEE Transactions on Software Engineering", March 2014, vol. 40, n^o 2, p. 123-136 [*DOI :* 10.1109/TSE.2013.50], https://hal.inria.fr/hal-01080198.

Publications of the year

Doctoral Dissertations and Habilitation Theses

[11] N. MITTAL.*Data, learning & privacy in recommandation systems*, Inria Rennes - Bretagne Atlantique ; ASAP, November 2016, https://hal.inria.fr/tel-01419886.

[12] J. OLIVARES.*Scaling out-of-core K-Nearest Neighbors Computation on Single Machines*, Inria Rennes, December 2016, https://hal.inria.fr/tel-01421362.

[13] A. RAULT.*User privacy in collaborative filtering systems*, Université Rennes 1, June 2016, https://tel.archives-ouvertes.fr/tel-01385503.

Articles in International Peer-Reviewed Journal

[14] M. BERTIER, M. PERRIN, C. TEDESCHI.*On the Complexity of Concurrent Multiset Rewriting*, in "International Journal of Foundations of Computer Science", 2016, vol. 27, n^o 1 [*DOI :* 10.1142/S0129054116500052], https://hal.inria.fr/hal-01326849.

[15] A. BOUTET, D. FREY, R. GUERRAOUI, A. JÉGOU, A.-M. KERMARREC.*Privacy-Preserving Distributed Collaborative Filtering*, in "Computing", August 2016, vol. 98, n^o 8, https://hal.inria.fr/hal-01251314.

[16] B. LE GAL, Y.-D. BROMBERG, L. RÉVEILLÈRE, J. SOLANKI.*A Flexible SoC and Its Methodology for Parser-Based Applications*, in "ACM Transactions on Reconfigurable Technology and Systems (TRETS)", 2016, vol. 10, n^o 1, 4 [*DOI :* 10.1145/2939379], https://hal.archives-ouvertes.fr/hal-01415653.

Invited Conferences

[17] M. RAYNAL.*A Look at Basics of Distributed Computing **, in "EEE ICDCS 2016 - 36th International Conference on Distributed Computing Systems", Nara, Japan, June 2016, https://hal.inria.fr/hal-01337523.

International Conferences with Proceedings

[18] T. ALLARD, D. FREY, G. GIAKKOUPIS, J. LEPILLER.*Lightweight Privacy-Preserving Averaging for the Internet of Things*, in "M4IOT 2016 - 3rd Workshop on Middleware for Context-Aware Applications in the IoT", Trento, Italy, ACM, December 2016, p. 19 - 22 [*DOI :* 10.1145/3008631.3008635], https://hal.inria.fr/hal-01421986.

[19] F. ANDRÉ, A.-M. KERMARREC, N. LE SCOUARNEC.*Cache locality is not enough: High-Performance Nearest Neighbor Search with Product Quantization Fast Scan*, in "42nd International Conference on Very Large Data Bases", New Delhi, India, September 2016, vol. 9, n^{o} 4, 12, https://hal.inria.fr/hal-01239055.

[20] O. BARAIS, J. BOURCIER, Y.-D. BROMBERG, C. DION.*Towards microservices architecture to transcode videos in the large at low costs*, in "TEMU 2016 - International Conference on Telecommunications and Multimedia", Heraklion, Greece, July 2016, p. 1 - 6 [*DOI :* 10.1109/TEMU.2016.7551918], https://hal.inria.fr/hal-01427277.

[21] Y. BENKAOUZ, M. ERRADI, A.-M. KERMARREC.*Nearest Neighbors Graph Construction: Peer Sampling to the Rescue*, in "4th International Conference, NETYS 2016, Marrakech, Morocco, May 18-20, 2016, Revised Selected Papers", Marrakech, Morocco, Networked Systems, Springer, May 2016, vol. 9944, p. 48 - 62 [*DOI :* 10.1007/978-3-319-46140-3_4], https://hal.inria.fr/hal-01407514.

[22] P. BERENBRINK, T. FRIEDETZKY, G. GIAKKOUPIS, P. KLING.*Efficient plurality consensus, or: The benefits of cleaning up from time to time*, in "43rd International Colloquium on Automata, Languages and Programming (ICALP 2016)", Rome, Italy, July 2016 [*DOI :* 10.4230/LIPIcs.ICALP.2016.136], https://hal.archives-ouvertes.fr/hal-01353690.

[23] P. BERENBRINK, G. GIAKKOUPIS, A.-M. KERMARREC, F. MALLMANN-TRENN.*Bounds on the Voter Model in Dynamic Networks*, in "ICALP 2016 - 43rd International Colloquium on Automata, Languages and Programming", Rome, Italy, July 2016 [*DOI :* 10.4230/LIPIcs.ICALP.2016.146], https://hal.archives-ouvertes.fr/hal-01353695.

[24] S. BOUGET.*Position paper: Toward an holistic approach of Systems of Systems*, in "Middleware 2016 – Doctoral Symposium", Trento, Italy, December 2016 [*DOI :* 10.1145/3009925.3009935], https://hal.inria.fr/hal-01419712.

[25] A. BOUTET, A.-M. KERMARREC, N. MITTAL, F. TAÏANI.*Being prepared in a sparse world: the case of KNN graph construction*, in "International Conference on Data Engineering (ICDE)", Helsinki, Finland, May 2016, https://hal.inria.fr/hal-01251010.

[26] A. CASTAÑEDA, C. DELPORTE-GALLET, H. FAUCONNIER, S. RAJSBAUM, M. RAYNAL.*Making Local Algorithms Wait-Free: The Case of Ring Coloring*, in "SSS", Lyon, France, Stabilization, Safety, and Security of Distributed Systems, November 2016, vol. 10083, 16, https://hal.inria.fr/hal-01416522.

[27] *Best Paper*
N. CHILUKA, A.-M. KERMARREC, J. OLIVARES.*The Out-of-core KNN Awakens: The light side of computation force on large datasets*, in "The International Conference on Networked Systems NETYS", Marrakech, Morocco, May 2016, https://hal.inria.fr/hal-01336673.

[28] C. DELPORTE-GALLET, H. FAUCONNIER, M. RAYNAL, S. RAJSBAUM.*Implementing Snapshot Objects on Top of Crash-Prone Asynchronous Message-Passing Systems*, in "Algorithms and Architectures for Parallel Processing (ICA3PP)", Granada, France, Algorithms and Architectures for Parallel Processing Volume 10048 of the series Lecture Notes in Computer Science pp 341-355, Springer, December 2016, vol. 10048, 15, https://hal.inria.fr/hal-01416498.

[29] C. DELPORTE-GALLET, H. FAUCONNIER, M. RAYNAL, S. RAJSBAUM.*t-Resilient Immediate Snapshot Is Impossible*, in "SIROCCO", Helsiinki, France, Structural Information and Communication Complexity, July 2016, vol. 9988, 15, https://hal.inria.fr/hal-01416515.

[30] *Best Paper*
S. DELBRUEL, D. FREY, F. TAÏANI.*Exploring The Use of Tags for Georeplicated Content Placement*, in "IEEE IC2E'16", Berlin, Germany, April 2016, https://hal.inria.fr/hal-01257939.

[31] S. DELBRUEL, D. FREY, F. TAÏANI.*Mignon: a Fast Decentralized Content Consumption Estimation in Large-Scale Distributed Systems*, in "16th IFIP International Conference on Distributed Applications and Interoperable Systems DAIS'16", Heraklion, Greece, June 2016, https://hal.inria.fr/hal-01301230.

[32] D. DIDONA, N. DIEGUES, R. GUERRAOUI, A.-M. KERMARREC, R. NEVES, P. ROMANO.*ProteusTM: Abstraction Meets Performance in Transactional Memory*, in "Twenty First International Conference on Architectural Support for Programming Languages and Operating Systems", Atlanta, United States, April 2016 [*DOI :* 10.1145/2872362.2872385], https://hal.inria.fr/hal-01250459.

[33] D. FREY, M. X. MAKKES, P.-L. ROMAN, F. TAÏANI, S. VOULGARIS.*Bringing secure Bitcoin transactions to your smartphone*, in "Proceedings of the 15th International Workshop on Adaptive and Reflective Middleware (ARM 2016)", Trento, Italy, ACM, December 2016, p. 3:1-3:6 [*DOI :* 10.1145/3008167.3008170], https://hal.inria.fr/hal-01384461.

[34] D. FREY, A. MOSTEFAOUI, M. PERRIN, P.-L. ROMAN, F. TAÏANI.*Speed for the elite, consistency for the masses: differentiating eventual consistency in large-scale distributed systems*, in "Proceedings of the 2016 IEEE 35th Symposium on Reliable Distributed Systems (SRDS 2016)", Budapest, Hungary, IEEE Computer Society, September 2016, p. 197-206 [*DOI :* 10.1109/SRDS.2016.30], https://hal.inria.fr/hal-01344138.

[35] G. GIAKKOUPIS, Y. NAZARI, P. WOELFEL.*How asynchrony affects rumor spreading time*, in "35th ACM Symposium on Principles of Distributed Computing (PODC 2016)", Chicago, United States, July 2016 [*DOI :* 10.1145/2933057.2933117], https://hal.archives-ouvertes.fr/hal-01353698.

[36] D. LIMA, H. MIRANDA, F. TAÏANI.*Partial Replication Policies for Dynamic Distributed Transactional Memory in Edge Clouds*, in "1st Workshop on Middleware for Edge Clouds & Cloudlets (MECC'16)", Trento, Italy, December 2016, p. 1 - 6 [*DOI :* 10.1145/3017116.3022872], https://hal.inria.fr/hal-01420708.

National Conferences with Proceeding

[37] J. L. RODAS, J. OLIVARES, J. A. GALINDO DUARTE, D. BENAVIDES.*Hacia el uso de sistemas de recomendación en sistemas de alta variabilidad*, in "CEDI 2016", Salamanca, Spain, September 2016, https://hal.inria.fr/hal-01342353.

Conferences without Proceedings

[38] R. ARIYATTU, F. TAÏANI.*Filament : a cohort construction service for decentralized collaborative editing platforms*, in "Compas 2016", Lorient, France, July 2016, https://hal.archives-ouvertes.fr/hal-01333513.

[39] H. LAKHLEF, D. FREY, M. RAYNAL.*Optimal Collision/Conflict-Free Distance-2 Coloring in Wireless Synchronous Broadcast/Receive Tree Networks*, in "45th International Conference on Parallel Processing",

Philadelphia, PA, United States, August 2016 [*DOI :* 10.1109/ICPP.2016.47], https://hal.archives-ouvertes.fr/hal-01396940.

Research Reports

[40] D. FREY, A. MOSTEFAOUI, M. PERRIN, A.-M. KERMARREC, C. MADDOCK, A. MAUTHE, P.-L. ROMAN, F. TAÏANI.*D.1.3 – Protocols for emergent localities*, IRISA ; LINA-University of Nantes ; Inria Rennes Bretagne Atlantique ; Lancaster University, July 2016, n^O D1.3, This report has been written by members of the consortium from IRISA and LINA partner, https://hal.archives-ouvertes.fr/hal-01343348.

[41] H. LAKHLEF, M. RAYNAL, F. TAÏANI.*Vertex Coloring with Communication and Local Memory Constraints in Synchronous Broadcast Networks*, IRISA ; Université de Rennes 1, 2016, n^O 2035, 23, https://hal.inria.fr/hal-01300095.

[42] A. MOSTEFAOUI, M. RAYNAL.*Time-Efficient Read/Write Register in Crash-prone Asynchronous Message-Passing Systems*, IRISA, January 2016, 14, https://hal.inria.fr/hal-01256067.

[43] M. RAYNAL.*Distributed Universal Constructions: a Guided Tour*, IRISA, November 2016, n^O 2040, 23, https://hal.inria.fr/hal-01397265.

Other Publications

[44] A. CASTAÑEDA, C. DELPORTE, H. FAUCONNIER, S. RAJSBAUM, M. RAYNAL.*Wait-freedom and Locality are not Incompatible (with Distributed Ring Coloring as an Example)*, January 2016, working paper or preprint, https://hal.inria.fr/hal-01265958.

[45] C. DELPORTE, H. FAUCONNIER, S. RAJSBAUM, M. RAYNAL.*Implementing Snapshot Objects on Top of Crash-Prone Asynchronous Message-Passing Systems*, May 2016, working paper or preprint, https://hal.inria.fr/hal-01313584.

[46] C. DELPORTE, H. FAUCONNIER, S. RAJSBAUM, M. RAYNAL.*t-Resilient Immediate Snapshot is Impossible*, May 2016, working paper or preprint, https://hal.inria.fr/hal-01315342.

[47] G. GIAKKOUPIS.*Amplifiers and Suppressors of Selection for the Moran Process on Undirected Graphs*, November 2016, working paper or preprint, https://hal.inria.fr/hal-01393044.

[48] A.-M. KERMARREC, N. MITTAL, J. OLIVARES.*Multithreading approach to process real-time updates in KNN algorithms*, December 2016, working paper or preprint, https://hal.inria.fr/hal-01415495.

[49] A. MOSTÉFAOUI, M. RAYNAL.*Two-Bit Messages are Sufficient to Implement Atomic Read/Write Registers in Crash-prone Systems*, February 2016, working paper or preprint, https://hal.inria.fr/hal-01271135.

[50] T. PETRIĆ, M. GOESSENS, L. NUAYMI, A. PELOV, L. TOUTAIN.*Measurements, Performance and Analysis of LoRa FABIAN, a real-world implementation of LPWAN*, June 2016, working paper or preprint, https://hal-institut-mines-telecom.archives-ouvertes.fr/hal-01331966.

References in notes

[51] M. AGUILERA.*A Pleasant Stroll Through the Land of Infinitely Many Creatures*, in "ACM SIGACT News, Distributed Computing Column", 2004, vol. 35, n^{o} 2.

[52] F. ANDRÉ.*Exploiting Modern Hardware for Large Scale Nearest Neighbor Search*, Université Rennes 1, November 2016.

[53] D. ANGLUIN.*Local and Global Properties in Networks of Processes*, in "Proc. 12th ACM Symposium on Theory of Computing (STOC'80)", 1980.

[54] K. BIRMAN, M. HAYDEN, O. OZKASAP, Z. XIAO, M. BUDIU, Y. MINSKY.*Bimodal Multicast*, in "ACM Transactions on Computer Systems", May 1999, vol. 17, n^{o} 2, p. 41-88.

[55] A. DEMERS, D. GREENE, C. HAUSER, W. IRISH, J. LARSON, S. SHENKER, H. STURGIS, D. SWINEHART, D. TERRY.*Epidemic algorithms for replicated database maintenance*, in "Proceedings of the ACM Symposium on Principles of Distributed Computing (PODC'87)", August 1987.

[56] P. EUGSTER, S. HANDURUKANDE, R. GUERRAOUI, A.-M. KERMARREC, P. KOUZNETSOV.*Lightweight Probabilistic Broadcast*, in "ACM Transaction on Computer Systems", November 2003, vol. 21, n^{o} 4.

[57] D. FREY, R. GUERRAOUI, A.-M. KERMARREC, A. RAULT.*Collaborative Filtering Under a Sybil Attack: Analysis of a Privacy Threat*, in "Eighth European Workshop on System Security EuroSec 2015", Bordeaux, France, April 2015 [*DOI :* 10.1145/2751323.2751328], https://hal.inria.fr/hal-01158723.

[58] M. MERRITT, G. TAUBENFELD.*Computing Using Infinitely Many Processes*, in "Proc. 14th Int'l Symposium on Distributed Computing (DISC'00)", 2000.

[59] S. RATNASAMY, P. FRANCIS, M. HANDLEY, R. KARP, S. SHENKER.*A Scalable Content-Addressable Network*, in "Conference of the Special Interest Group on Data Communication (SIGCOMM'01)", 2001.

[60] A. ROWSTRON, P. DRUSCHEL.*Pastry: Scalable, distributed object location and routing for large-scale peer-to-peer systems*, in "IFIP/ACM Intl. Conf. on Distributed Systems Platforms (Middleware)", 2001.

[61] I. STOICA, R. MORRIS, D. KARGER, F. KAASHOEK, H. BALAKRISHNAN.*Chord: A Scalable Peer-to-peer Lookup Service for Internet Applications*, in "SIGCOMM'01", 2001.

[62] S. VOULGARIS, D. GAVIDIA, M. VAN STEEN.*CYCLON: Inexpensive Membership Management for Unstructured P2P Overlays*, in "Journal of Network and Systems Management", 2005, vol. 13, n^{o} 2.

Project-Team ASCOLA

Aspect and composition languages

IN COLLABORATION WITH: Laboratoire des Sciences du numerique de Nantes

IN PARTNERSHIP WITH:
CNRS

Ecole des Mines de Nantes

IMT Atlantique Bretagne-Pays de la Loire

Université Nantes

RESEARCH CENTER
Rennes - Bretagne-Atlantique

THEME
Distributed programming and Software engineering

Table of contents

Project-Team ASCOLA

Creation of the Project-Team: 2009 January 01

Keywords:

Computer Science and Digital Science:

1.1.6. - Cloud
1.1.8. - Security of architectures
1.1.13. - Virtualization
1.3. - Distributed Systems
1.6. - Green Computing
2.1. - Programming Languages
2.1.1. - Semantics of programming languages
2.1.2. - Object-oriented programming
2.1.3. - Functional programming
2.1.4. - Aspect-oriented programming
2.1.6. - Concurrent programming
2.1.7. - Distributed programming
2.1.10. - Domain-specific languages
2.1.11. - Proof languages
2.2.1. - Static analysis
2.4.2. - Model-checking
2.4.3. - Proofs
2.5. - Software engineering
2.6.2. - Middleware
2.6.3. - Virtual machines
3.1.3. - Distributed data
3.1.5. - Control access, privacy
3.1.8. - Big data (production, storage, transfer)
4.5. - Formal methods for security
4.6. - Authentication
4.7. - Access control
4.8. - Privacy-enhancing technologies
7.1. - Parallel and distributed algorithms
7.4. - Logic in Computer Science

Other Research Topics and Application Domains:

3.1. - Sustainable development
4.5. - Energy consumption
4.5.1. - Green computing
5.1. - Factory of the future
6.1. - Software industry
6.1.1. - Software engineering
6.1.2. - Software evolution, maintenance
6.5. - Information systems

1. Members

Research Scientists

Adrien Lebre [Inria, Researcher, on leave from MN]
Guillaume Munch [Inria, Researcher, from Oct 2016]
Nicolas Tabareau [Inria, Researcher]

Faculty Members

Mario Südholt [Team leader, MN, Professor, HDR]
Frederico Alvares de Oliveira Junior [MN, Associate Professor, temporary position]
Julien Cohen [Univ. Nantes, Associate Professor]
Pierre Cointe [MN, Professor, HDR]
Hélène Coullon [MN, Associate Professor, Inria chair, from Oct 2016]
Rémi Douence [MN, Associate Professor, HDR]
Hervé Grall [MN, Associate Professor]
Thomas Ledoux [MN, Associate Professor]
Jean-Marc Menaud [MN, Professor, HDR]
Jacques Noyé [MN, Associate Professor]
Jean-Claude Royer [MN, Professor, HDR]

Technical Staff

Ronan-Alexandre Cherrueau [Inria, from Oct 2016]
Rémy Pottier [MN]

PhD Students

Mohamed Abderrahim [Orange Labs, granted by CIFRE]
Mohammad-Mahdi Bazm [Orange Labs, from Jul 2016]
Walid Benghabrit [ARMINES, until Sep 2016]
Paul Blouët [MN]
Simon Boulier [MN, (ASN from ENS Rennes)]
Bastien Confais [CNRS, from Oct 2016]
Ismael Cuadrado Cordero [MN]
Frédéric Dumont [EasyVirt, until Jun 2016, granted by CIFRE]
Simon Dupont [Sigma, until Apr 2016, granted by CIFRE]
Alexandre Garnier [MN, until Sep 2016]
Gaetan Gilbert [MN (ASN from ENS Lyon), from Sept 2016]
Md Sabbir Hasan [Inria, co-supervision with Prof. Pazat, Myriads team, Inria]
Yacine Hebbal [Orange Labs, granted by CIFRE]
Ambroise Lafont [MN (X Grant), from Oct 2016]
Gabriel Lewertowski [Inria, until Oct 2016]
Florent Marchand de Kerchove [MN, until Apr 2016]
Thuy Linh Nguyen [Inria]
Jonathan Pastor [MN, until March 2016]
Kevin Quirin [MN]

Post-Doctoral Fellows

Benedikt Ahrens [Inria, from May 2016]
Zakarea Al Shara [MN, from Oct 2016]
Ali Kassem [Inria]
Guillaume Le Louet [MN, until Jul 2016]
Pierre-Marie Pedrot [Inria, until Sep 2016]
Dimitri Pertin [Inria, from Oct 2016]
Anthony Simonet [Inria]

Visiting Scientists

Paige North [Inria, from Sep 2016]
Egbert Rijke [Carnegie Mellon Uni., from Jun 2016 until Jul 2016]

Administrative Assistants

Anne Claire Binétruy [Inria, (part time 30%)]
Florence Rogues [MN, (part time 30%)]

Others

Marie Delavergne [Inria, Master student, from Apr 2016 until Aug 2016]
Theo Winterhalter [ENS Cachan, Master student, from Mar 2016 until Aug 2016]

2. Overall Objectives

2.1. Presentation

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

We pursue the following objectives:

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

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

3. Research Program

3.1. Overview

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

3.2. Software Composition

Modules and services. The idea that building *software components*, i.e., composable prefabricated and parameterized software parts, was key to create an effective software industry was realized very early [72]. At that time, the scope of a component was limited to a single procedure. In the seventies, the growing complexity of software made it necessary to consider a new level of structuring and programming and led to the notions of information hiding, *modules*, and module interconnection languages [79], [55]. Information hiding promotes a black-box model of program development whereby a module implementation, basically a collection of procedures, is strongly encapsulated behind an interface. This makes it possible to guarantee logical invariant *properties* of the data managed by the procedures and, more generally, makes *modular reasoning* possible.

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

Object-Oriented Programming. Classes and objects provide another kind of software component, which makes it necessary to distinguish between *component types* (classes) and *component instances* (objects). Indeed, unlike modules, objects can be created dynamically. Although it is also possible to talk about classes in terms of interfaces and implementations, the encapsulation provided by classes is not as strong as the one provided by modules. This is because, through the use of inheritance, object-oriented languages put the emphasis on *incremental programming* to the detriment of modular programming. This introduces a white-box model of software development and more flexibility is traded for safety as demonstrated by the *fragile base class* issue [75].

Architecture Description Languages. The advent of distributed applications made it necessary to consider more sophisticated connections between the various building blocks of a system. The *software architecture* [84] of a software system describes the system as a composition of *components* and *connectors*, where the connectors capture the *interaction protocols* between the components [43]. It also describes the rationale behind such a given architecture, linking the properties required from the system to its implementation. *Architecture Description Languages* (ADLs) are languages that support architecture-based development [73]. A number of these languages make it possible to generate executable systems from architectural descriptions, provided implementations for the primitive components are available. However, guaranteeing that the implementation conforms to the architecture is an issue.

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

Object interactions [77], component interactions [90], [81] and service orchestrations [56] are most frequently expressed in terms of *regular interaction protocols* that enable basic properties, such as compatibility, substitutability, and deadlocks between components to be defined in terms of basic operations and closure properties of finite-state automata. Furthermore, such properties may be analyzed automatically using, e.g., model checking techniques [53], [62].

However, the limited expressive power of regular languages has led to a number of approaches using more expressive *non-regular* interaction protocols that often provide distribution-specific abstractions, e.g., session types [66], or context-free or turing-complete expressiveness [82], [50]. While these protocol types allow conformance between components to be defined (e.g., using unbounded counters), property verification can only be performed manually or semi-automatically.

3.3. Programming languages for advanced modularization

The main driving force for the structuring means, such as components and modules, is the quest for clean *separation of concerns* [57] on the architectural and programming levels. It has, however, early been noted that concern separation in the presence of crosscutting functionalities requires specific language and implementation level support. Techniques of so-called *computational reflection*, for instance, Smith's 3-Lisp or Kiczales's CLOS meta-object protocol [85], [69] as well as metaprogramming techniques have been developed to cope with this problem but proven unwieldy to use and not amenable to formalization and property analysis due to their generality. Methods and techniques from two fields have been particularly useful in addressing such advanced modularization problems: Aspect-Oriented Software Development as the field concerned with the systematic handling of modularization issues and domain-specific languages that provide declarative and efficient means for the definition of crosscutting functionalities.

Aspect-Oriented Software Development [68], [41] has emerged over the previous decade as the domain of systematic exploration of crosscutting concerns and corresponding support throughout the software development process. The corresponding research efforts have resulted, in particular, in the recognition of *crosscutting* as a fundamental problem of virtually any large-scale application, and the definition and implementation of a large number of aspect-oriented models and languages.

However, most current aspect-oriented models, notably AspectJ [67], rely on pointcuts and advice defined in terms of individual execution events. These models are subject to serious limitations concerning the modularization of crosscutting functionalities in distributed applications, the integration of aspects with other modularization mechanisms such as components, and the provision of correctness guarantees of the resulting AO applications. They do, in particular, only permit the manipulation of distributed applications on a per-host basis, that is, without direct expression of coordination properties relating different distributed entities [86]. Similarly, current approaches for the integration of aspects and (distributed) components do not directly express interaction properties between sets of components but rather seemingly unrelated modifications to individual components [54]. Finally, current formalizations of such aspect models are formulated in terms of low-level semantic abstractions (see, e.g., Wand's et al semantics for AspectJ [89]) and provide only limited support for the analysis of fundamental aspect properties.

Different approaches have been put forward to tackle these problems, in particular, in the context of so-called *stateful* or *history-based aspect languages* [58], [59], which provide pointcut and advice languages that directly express rich relationships between execution events. Such languages have been proposed to directly express coordination and synchronization issues of distributed and concurrent applications [78], [48], [61], provide more concise formal semantics for aspects and enable analysis of their properties [44], [60], [58], [42]. Furthermore, first approaches for the definition of *aspects over protocols* have been proposed, as well as over regular structures [58] and non-regular ones [88], [76], which are helpful for the modular definition and verification of protocols over crosscutting functionalities.

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

Domain-specific languages (DSLs) represent domain knowledge in terms of suitable basic language constructs and their compositions at the language level. By trading generality for abstraction, they enable complex relationships among domain concepts to be expressed concisely and their properties to be expressed and formally analyzed. DSLs have been applied to a large number of domains; they have been particularly popular in the domain of software generation and maintenance [74], [92].

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

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

3.4. Distribution and Concurrency

While ASCOLA does not investigate distribution and concurrency as research domains per se (but rather from a software engineering and modularization viewpoint), there are several specific problems and corresponding approaches in these domains that are directly related to its core interests that include the structuring and modularization of large-scale distributed infrastructures and applications. These problems include crosscutting functionalities of distributed and concurrent systems, support for the evolution of distributed software systems, and correctness guarantees for the resulting software systems.

Underlying our interest in these domains is the well-known observation that large-scale distributed applications are subject to *numerous crosscutting functionalities* (such as the transactional behavior in enterprise information systems, the implementation of security policies, and fault recovery strategies). These functionalities are typically partially encapsulated in distributed infrastructures and partially handled in an ad hoc manner by using infrastructure services at the application level. Support for a more principled approach to the development and evolution of distributed software systems in the presence of crosscutting functionalities has been investigated in the field of *open adaptable middleware* [49], [71]. Open middleware design exploits the concept of reflection to provide the desired level of configurability and openness. However, these approaches are subject to several fundamental problems. One important problem is their insufficient, framework-based support that only allows partial modularization of crosscutting functionalities.

There has been some *criticism* on the use of *AspectJ-like aspect models* (which middleware aspect models like that of JBoss AOP are an instance of) for the modularization of distribution and concurrency related concerns, in particular, for transaction concerns [70] and the modularization of the distribution concern itself [86]. Both criticisms are essentially grounded in AspectJ's inability to explicitly represent sophisticated relationships between execution events in a distributed system: such aspects therefore cannot capture the semantic relationships that are essential for the corresponding concerns. History-based aspects, as those proposed by the ASCOLA project-team provide a starting point that is not subject to this problem.

From a point of view of language design and implementation, aspect languages, as well as domain specific languages for distributed and concurrent environments share many characteristics with existing distributed languages: for instance, event monitoring is fundamental for pointcut matching, different synchronization strategies and strategies for code mobility [64] may be used in actions triggered by pointcuts. However, these relationships have only been explored to a small degree. Similarly, the formal semantics and formal properties of aspect languages have not been studied yet for the distributed case and only rudimentarily for the concurrent one [44], [61].

3.5. Security

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

Two main issues in this space are particularly problematic from a compositional point of view. First, information flow properties of programming languages, such as flow properties of Javascript [46], and service-based systems [52] are typically specially-tailored to specific properties, as well as difficult to express and analyze. Second, the enforcement of security properties and security policies, especially accountability-related properties [80], [87], is only supported using ad hoc means with rudimentary support for property verification.

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

3.6. Green IT

With the emergence of the Future Internet and the dawn of new IT architecture and computation models such as cloud computing, the usage of data centers (DC) as well as their power consumption increase dramatically [51]. Besides the ecological impact [65], energy consumption is a predominant criterion for DC providers since it determines the daily cost of their infrastructure. As a consequence, power management becomes one of the main challenges for DC infrastructures and more generally for large-scale distributed systems.

To address this problem, we study two approaches: a workload-driven [47] and power-driven one [83]. As part of the workload-driven solution, we adapt the power consumption of the DC depending on the application workload, and evaluate whether this workload to be more reactive. We develop a distributed system from the system to the service-oriented level mainly based on hardware and virtualization capabilities that is managed in a user-transparent fashion. As part of the power-driven approach, we address energy consumption issues through a strong synergy inside the infrastructure software stack and more precisely between applications and resource management systems. This approach is characterized by adapting QoS properties aiming at the best trade-off between cost of energy (typically from the regular electric grid), its availability (for instance, from renewable energy), and service degradation caused, for instance, by application reconfigurations to jobs suspensions.

3.7. Capacity Planning for Large Scale Distributed System

Since the last decade, cloud computing has emerged as both a new economic model for software (provision) and as flexible tools for the management of computing capacity [45]. Nowadays, the major cloud features have become part of the mainstream (virtualization, storage and software image management) and the big market players offer effective cloud-based solutions for resource pooling. It is now possible to deploy virtual infrastructures that involve virtual machines (VMs), middleware, applications, and networks in such a simple manner that a new problem has emerged since 2010: VM sprawl (virtual machine proliferation) that consumes valuable computing, memory, storage and energy resources, thus menacing serious resource shortages. Scientific approaches that address VM sprawl are both based on classical administration techniques like the lifecycle management of a large number of VMs as well as the arbitration and the careful management of all resources consumed and provided by the hosting infrastructure (energy, power, computing, memory, network etc.) [63], [91].

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

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

4. Application Domains

4.1. Enterprise Information Systems and Services

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

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

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

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

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

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

In 2015, we have published results on constructive mechanisms for security and accountability properties in service-based systems as well as results on service provisioning problems, in particular, service interoperability and mediation. Furthermore, we take part in the European project A4Cloud on accountability challenges, that is, the responsible stewardship of third-party data and computations, see Sec. 9.3.

4.2. Capacity Planning in Cloud, Fog and Edge Computing

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

In 2015, we have proposed VMPlaces, a dedicated framework to evaluate and compare VM placement algorithms. Globally the framework is composed of two major components: the injector and the VM placement algorithm. The injector constitutes the generic part of the framework (i.e. the one you can directly use) while the VM placement algorithm is the component a user wants to study (or compare with other existing algorithms), see Sec. 7.2.

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

4.3. Pervasive Systems

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

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

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

In 2014, we have extended the language EScala, which integrates reactive programming through events with aspect-oriented and object-oriented mechanisms.

5. Highlights of the Year

5.1. Highlights of the Year

This year the team has produced major results in the domains of the foundations of computer science as well as capacity management for large-scale distributed software systems.

Concerning the foundations of computer science, we have presented new results on the provably correct execution of programs that are only partially typed [22] and generalized the use of dependent types with side effects [26].

As to distributed systems, we have introduced a new cloud model that provides QoS-levels and SLA as first-class citizens of cloud-based systems [19]. Furthermore, we have provided new mechanisms for the privacy-preserving storage of data of a user over clouds managed by different cloud providers [30].

6. New Software and Platforms

6.1. CSLA

Cloud Service Level Agreement language

KEYWORDS: Cloud computing - Service-level agreement - Elasticity

FUNCTIONAL DESCRIPTION

CSLA, the Cloud Service Level Agreement language, allows the definition of SLA properties for arbitrary Cloud services (XaaS). CSLA addresses QoS uncertainty in unpredictable and dynamic environment and provides a cost model of Cloud computing. Besides the standard formal definition of contracts – comprising validity, parties, services definition and guarantees/violations – CSLA is enriched with features, such as QoS degradation and an advanced penalty model, thus introducing fine-grained language support for Cloud elasticity management.

- Participants: Thomas Ledoux and Yousri Kouki
- Contact: Thomas Ledoux
- URL: http://www.emn.fr/z-info/csla/

6.2. CSQL

Cryptographic Composition for Query Language

SCIENTIFIC DESCRIPTION

C2QL is a compositional language of security techniques for information privacy in the cloud. A cloud service can use security techniques to ensure information privacy. These techniques protect privacy by converting the client's personal data into unintelligible text. But they also cause the loss of some functionalities of the service. As a solution, CSQL permits to compose security techniques to ensure information privacy without the loss of functionalities. But, the composition makes the writing of programs more intricate. To help the programmer, C2QL defines a query language for the definition of cloud services that enforces information privacy with the composition of security techniques. This language comes with a set of algebraic laws to, systematically, transform a local service without protection into its cloud equivalent that is protected by composition.

FUNCTIONAL DESCRIPTION

C2QL is implemented in Idris, a functional language of the Haskell family. The implementation harnesses the Idris dependant type system to ensure the correct composition of security mechanisms and provides a transformation of the implementation into a π-calculus. This transformation serves two purposes. First, it makes the distribution explicit, showing how a computation is distributed over SaaS, PaaS and client applications. Then, it helps defining an encoding into ProVerif to check that the service preserves the privacy of its clients.

- Participants: Ronan-Alexandre Cherrueau, Rémi Douence, Mario Südholt
- Contact: Ronan-Alexandre Cherrueau
- URL: https://github.com/rcherrueau/C2QL

6.3. EScala

SCIENTIFIC DESCRIPTION

EScala extends the idea of events as object members, as realized by C# events, with the possibility to define events declaratively by expressions over other events. The occurrences of an event can be defined by various set operations, such as union, intersection and difference, applied on the occurrences of other events. Events can be filtered by arbitrary conditions, the data attached to the events can be transformed by arbitrary functions. Event expressions make it possible to define events in terms of other events, at the lowest level relying on primitive events.

FUNCTIONAL DESCRIPTION

EScala is an extension of Scala programming language with support for events as attributes of objects. The support for events in EScala, combine the ideas of event-driven, aspect-oriented and functional-reactive programming.

- Participants: Jacques Noyé and Jurgen Van Ham
- Contact: Jurgen Van Ham
- URL: http://www.stg.tu-darmstadt.de/research/escala/index.en.jsp

6.4. JEScala

FUNCTIONAL DESCRIPTION

JEScala extends EScala with support for concurrent programming. Events can be declared as asynchronous so that their handling takes place concurrently. A new composition operator, the join operator, inspired by the join calculus, can also be used to synchronize the concurrent activities created by asynchronous events and communicate between them.

- Participants: Jurgen Van Ham and Jacques Noyé
- Contact: Jurgen Van Ham
- URL: http://www.stg.tu-darmstadt.de/research/jescala_menu/index.en.jsp

6.5. SimGrid

Scientific Instrument for the study of Large-Scale Distributed Systems
KEYWORDS: Large-scale Emulators - Grid Computing - Distributed Applications
FUNCTIONAL DESCRIPTION

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

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

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

- Participants: Frederic Suter, Martin Quinson, Arnaud Legrand, Takahiro Hirofuchi, Adrien Lebre, Jonathan Pastor, Mario Sudholt, Luka Stanisic, Augustin Degomme, Jean Marc Vincent, Florence Perronnin and Jonathan Rouzaud-Cornabas
- Partners: CNRS - ENS Rennes - Université de Nancy
- Contact: Martin Quinson
- URL: http://simgrid.gforge.inria.fr/

6.6. VMPlaces

FUNCTIONAL DESCRIPTION

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

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

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

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

- Participants: Takahiro Hirofuchi, Adrien Lebre, Jonathan Pastor, Flavien Quesnel and Mario Sudholt
- Contact: Adrien Lebre
- URL: http://beyondtheclouds.github.io/VMPlaceS/

6.7. btrCloud

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

FUNCTIONAL DESCRIPTION

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

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

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

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

Finally, several kinds of optimizations are currently available, such as energy and load balancing. The former can help save up to around 20% of the data-center energy consumption. The latter provides optimized quality of service properties for applications that are hosted in the virtualized datacenters.

- Participants: Guillaume Le Louet, Frederic Dumont and Jean-Marc Menaud
- Contact: Jean-Marc Menaud
- URL: http://www.btrcloud.org/btrCloud/index_EN.html

7. New Results

7.1. Software composition and programming languages

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

This year we have published a number of new results in the domains of software composition and programming languages that range from pragmatic ones like modularity issues to formal studies in the domain of dependent type theory via static analysis and formal verification.

7.1.1. Formal Methods, logics and type theory

Concerning verification and formal semantics, we have defined the semantics of our dependent interoperability framework and we propose the notion the partial type equivalences as a key feature. We have also studied proofs in dependent type theory and synthesized call-by-value and call-by-name translations.

7.1.1.1. Verified Dependent Interoperability.

Full-spectrum dependent types promise to enable the development of correct-by-construction software. However, even certified software needs to interact with simply-typed or untyped programs, be it to perform system calls, or to use legacy libraries. Trading static guarantees for runtime checks, the dependent interoperability framework provides a mechanism by which simply-typed values can safely be coerced to dependent types and, conversely, dependently-typed programs can defensively be exported to a simply-typed application. In [22], we give a semantic account of dependent interoperability. Our presentation relies on and is guided by a pervading notion of type equivalence, whose importance has been emphasized in recent works on homotopy type theory. Specifically, we develop the notion of partial type equivalences as a key foundation for dependent interoperability. Our framework is developed in Coq; it is thus constructive and verified in the strictest sense of the terms. Using our library, users can specify domain-specific partial equivalences between data structures. Our library then takes care of the (sometimes, heavy) lifting that leads to interoperable programs. It thus becomes possible, as we shall illustrate, to internalize and hand-tune the extraction of dependently-typed programs to interoperable OCaml programs within Coq itself.

7.1.1.2. Forcing in Type Theory.

In [26], we study forcing translations of proofs in dependent type theory, through the Curry-Howard correspondence. Based on a call-by-push-value decomposition, we synthesize two simply-typed translations: i) one call-by-value, corresponding to the translation derived from the presheaf construction as studied in a previous paper; ii) one call-by-name, whose intuitions already appear in Krivine and Miquel's work. Focusing on the call-by-name translation, we adapt it to the dependent case and prove that it is compatible with the definitional equality of our system, thus avoiding coherence problems. This allows us to use any category as forcing conditions, which is out of reach with the call-by-value translation. Our construction also exploits the notion of storage operators in order to interpret dependent elimination for inductive types. This is a novel example of a dependent theory with side-effects, clarifying how dependent elimination for inductive types must be restricted in a non-pure setting. Being implemented as a Coq plugin, this work gives the possibility to formalize easily consistency results, for instance the consistency of the negation of Voevodsky's univalence axiom.

7.1.2. Programming languages

In the domain of programming languages we have presented new results on constraint programming, development of correct programs by construction and better controls for computational effects and modularity for JavaScript.

7.1.2.1. Constraint programming

Constraint programming (CP) relies on filtering algorithms in order to deal with combinatorial problems. Global constraints offer efficient algorithms for complex constraints. In particular a large family of global constraints can be expressed as constraints of finite state automata with counters. We have generalized these automata constraints in order to compose them as transducers [16]. We have also extended these results with different techniques [20]. First, we have improved the automaton synthesis to generate automata with fewer accumulators. Second, we have shown how to decompose a constraint specified by an automaton with accumulators into a conjunction of linear inequalities, for use by a MIP (Mixed-Integer Programming) solver. Third, we have generalized the implied constraint generation to cover the entire family of time-series constraints. The newly synthesized automata for time-series constraints outperform the old ones, for both the CP and MIP decompositions, and the generated implied constraints boost the inference, again for both the CP and MIP decompositions.

7.1.2.2. Program correctness

Most IDEs provide refactoring tools to assist programmers when they modify the structure of their software. However the refactoring facilities of many popular tools (Eclipse, Visual Studio, IntelliJ, etc.) are currently not reliable : they occasionally change the program semantics in unexpected ways, and, as a result, the programmers systematically have to re-test the resulting code. We have build a refactoring tool for C programs which core operation is proved correct by construction [21]. To do that, we build an AST transformation with Coq (based on the CompCert C implementation) and we prove that this transformation preserves the external behavior of programs. The code of the transformation is then extracted to OCaml and is then embedded in a traditional parse/transform/pretty-print setting to provide a working prototype.

7.1.2.3. Effect Capabilities

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 propose in [18] effect capabilities 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 further refine capabilities with type-based permission lattices to allow fine-grained decomposition of authority. We provide an implementation of effect capabilities in Haskell, 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. We first exemplify how to tame effect interferences using effect capabilities by treating state and exceptions. Then we focus on taming I/O by proposing a fine-grained lattice of I/O permissions based on the current classification of its operations. Finally, we show that integrating effect capabilities with modern tag-based monadic mechanisms provides a practical, modular and safe mechanism for monadic programming in Haskell.

7.1.2.4. Extensible JavaScript Modules

As part of the SecCloud project, we have studied how to modularly extend JavaScript interpreters with dynamic security analyses in particular information flow analyses. This has led us to study ways to improve on the standard JavaScript module pattern. This pattern is commonly used to encapsulate definitions by using closures. However, closures prevent module definitions from being extended at runtime. We have proposed a simple pattern that not only opens the module, but allows one to extend the module definitions in layers [39]. The pattern leverages the with construct and the prototype delegation mechanism of JavaScript to mimic a form of dynamic binding, while minimizing the changes made to the module code.

Florent Marchand's PhD thesis [13] details the proposal further and shows its application to the modular extension of Narcissus, a full-blown JavaScript interpreter, with several dynamic analyses, including the information flow of Austin and Flanagan based on multiple facets. A comparison with a previous ad hoc implementation of the analysis illustrates the benefits of the proposal.

7.1.3. Software Security and Privacy

In the area of security we have focused on expressing advanced security concerns with abstract and formal languages and the study of policy monitoring and the detection of conflicts.

7.1.3.1. Runtime verification of advanced logical security properties.

Monitoring or runtime verification means to observe the system execution and to check if it deviates or not from a predefined contract. Our contract is a formula written in AAL (Abstract Accountability Language) expressing the expected behavior of a system, the audit steps as well as punishment and compensation. We choose to use the rewriting approach with the three valued logic as many other existing approaches. The monitoring problem raised a validity question, if we start with a formula neither true nor false are we sure to conclude? The response is no and this is a completeness problem and all published solutions are incomplete. For LTL, mixing the standard semantics, the rewriting principle and coinduction we are able to define a complete monitoring mechanism. A first implementation has been done into our AccLab tool support and sketched in [38]. We are investigating the extension of our LTL rewriting mechanism to cope with the first-order case.

7.1.3.2. Specification of advanced security and privacy properties.

Security and privacy requirements in ubiquitous systems need a sophisticated policy language with features to express access restrictions and obligations. Ubiquitous systems involve multiple actors owning sensitive data concerning aspects such as location, discrete and continuous time, multiple roles that can be shared among actors or evolve over time. Conflict management is an important problem in security policy frameworks. In [31] we present an abstract language (AAL) dedicated to accountability. We show how to specify most of these security and privacy features and compare it with the XACML approach. We also classified the existing conflict detection for XACML like approaches in dynamic, testing, or static detection. A thorough analysis of these mechanisms reveals that they have several weaknesses and they are not applicable in our context. We advocate for a classic approach using the notion of logical consistency to detect conflicts in AAL.

7.1.3.3. Composition of privacy-enhancing and security mechanisms.

As part of his PhD thesis [11], Ronan Cherrueau's has defined a language for the composition of three privacy-enhancing and security mechanisms: symmetric key encryption, database fragmentation and on-client computations. The language allows the expression of distributed programs that protect data by applying compositions of the three mechanisms to them. The language ensures basic privacy and security properties by a type system based on dependent types. This type system ensures, for example, that data that has been encrypted and stored in a database fragment cannot be accessed in plain form and from another location than that fragment. Furthermore, the language comes equiped with four major additional results. First, a calculus that allows for the semi-automatic derivation of distributed privacy-preserving and secure programs from an original non-distributed one. Second, a transformation from the language to the π-calculus. Third, a transformation into an input specification to the Proverif model checker for security properties. Fourth, two implementations on the basis of, respectively, the Scala and Idris languages that harness their corresponding dependent type systems.

7.2. Distributed programming and the Cloud

Participants: Frederico Alvares, Bastien Confais, Simon Dupont, Md Sabbir Hasan, Adrien Lebre, Thomas Ledoux, Guillaume Le Louët, Jean-Marc Menaud, Jonathan Pastor, Rémy Pottier, Anthony Simonet, Mario Südholt.

7.2.1. Cloud applications and infrastructures

Complex event processing. We presented this year the evolution of SensorScript towards a language for complex event processing dedicated to sensor networks. While the model mainly relies on previous works, we highlighted how the new language builds on the multitree in order to provide complex event processing mechanisms. We are able to balance the syntactic concision of the language with a real-time complex event processor for sensor networks. By providing flexible selections over the nodes, with the possibility to filter them on complex conditions, possibly over a time window, we offer a strong alternative to traditional SQL used in the literature. Moreover, SensorScript does not focus only on data access. In fact it provides the possibility to widen the scope of the methods accessible on nodes to other features than sensors monitoring, including but not limited to addressing actuators functions. Finally we showed that SensorScript is able to address examples proposed in the literature, with simpler results than SQL, while highlighting its limitations, especially on history management. [24]

Secure cloud storage. The increasing number of cloud storage services like Dropbox or Google Drive allows users to store more and more data on the Internet. However, these services do not give users enough guarantees in protecting the privacy of their data. In order to limit the risk that the storage service scans user documents for commercial purposes, we propose a storage service that stores data on several cloud providers while preventing these providers to read user documents. TrustyDrive is a cloud storage service that protects the privacy of users by breaking user documents into blocks in order to spread them on several cloud providers. As cloud providers only own a part of the blocks and they do not know the block organization, they can not read user documents. Moreover, the storage service connects directly users and cloud providers without using a third-party as is generally the practice in cloud storage services. Consequently, users do not give critical information (security keys, passwords, etc.) to a third-party. [30]

7.2.1.1. Service-level agreement for the Cloud.

Quality-of-service and SLA guarantees are among the major challenges of cloud-based services. In [19], we first present a new cloud model called SLAaaS — SLA aware Service. SLAaaS considers QoS levels and SLA as first class citizens of cloud-based services. This model is orthogonal to other SaaS, PaaS, and IaaS cloud models, and may apply to any of them. More specifically, we make three contributions: (i) we provide a domain-specific language that allows to define SLA constraints in cloud services; (ii) we present a general control-theoretic approach for managing cloud service SLA; (iii) we apply our approach to MapReduce, locking, and e-commerce services.

7.2.1.2. Cloud Capacity Planning and Elasticity.

Capacity management is a process used to manage the capacity of IT services and the IT infrastructure. Its primary goal is to ensure that IT resources (services, infrastructure) are right-sized to meet current and future requirements in a cost-effective and timely manner. In [34], we present a comprehensive overview of capacity planning and management for cloud computing. First, we state the problem of capacity management in the context of cloud computing from the point of view of several service providers. Second, we provide a brief discussion about *when* capacity planning should take place. Finally, we survey a number of methods for capacity planning and management proposed by both people from industry and researchers.

In his PhD [12], Simon Dupont proposes to extend the concept of elasticity to higher layers of the cloud, and more precisely to the SaaS level. He presents the new concept of *software elasticity* by defining the ability of the software to adapt, ideally in an autonomous way, to cope with workload changes and/or limitations of IaaS elasticity. This brings the consideration of Cloud elasticity in a multi-layer way through the adaptation of all kind of Cloud resources (software, virtual machines, physical machines). In [23], we introduce ElaScript, a DSL that offers Cloud administrators a simple and concise way to define complex elasticity-based reconfiguration plans. ElaScript is capable of dealing with both infrastructure and software elasticities, independently or together, in a coordinated way. We validate our approach by first showing the interest to have a DSL offering multiple levels of control for Cloud elasticity, and then by showing its integration with a realistic well-known application benchmark deployed in OpenStack and Grid'5000 infrastructure testbed.

7.2.1.3. Infrastructure.

Academic and industry experts are now advocating for going from large-centralized Cloud Computing infrastructures to smaller ones massively distributed at the edge of the network (aka., Fog and Edge Computing solutions). Among the obstacles to the adoption of this model is the development of a convenient and powerful IaaS system capable of managing a significant number of remote data-centers in a unified way.

In 2016, we achieved three major results in this context.

The first result is related to the economical viability of Fog/Edge Computing infrastructures that is often debated w-r-t large cloud computing data centers operated by US giants such as Amazon, Google .... To answer such a question, we conducted a specific study that goes beyond the state of the art of the current cost model of Distributed Cloud infrastructures. First, we provided a classification of the different ways of deploying Distributed Cloud platforms. Then, we proposed a versatile cost model that can help new actors evaluate the viability of deploying a Fog/Edge Computing offer. We illustrated the relevance of our proposal by instantiating it over three use-cases and comparing them according to similar computation capabilities provided by the Amazon solution. Such a study clearly showed that deploying a Distributed Cloud infrastructure makes sense for telcos as well as new actors willing to enter the game [29].

The second result is related to the preliminary revisions we made in OpenStack. The OpenStack software suite has become the de facto open-source solution to operate, supervise and use a Cloud Computing infrastructure. Our objective is to study to what extent current OpenStack mechanisms can handle massively distributed cloud infrastructures and to propose revisions/extensions of internal mechanisms when appropriate. The work we conducted this year focused on the Nova service of OpenStack.More precisely, we modified the code base in order to use a distributed key/value store instead of the centralized SQL backend. We conducted several experiments that validate the correct behavior and gives performance trends of our prototype through an emulation of several data-centers using Grid'5000 testbed. In addition to paving the way to the first large-scale and Internet-wide IaaS manager, we expect this work will attract a community of specialists from both distributed system and network areas to address the Fog/Edge Computing challenges within the OpenStack ecosystem [36], [27]. These and additional corresponding results have been presented in a more detailed manner as part of Jonathan Pastor's PhD thesis [14].

The third result is related to the data management in Fog/Edge Computing infrastructures. Our ultimate goal is to propose an Amazon-S3 like system, *i.e.*, a blob storage service, that can take into account Fog/Edge specifics. The study we achieved this year is preliminary. We first identified a list of properties a storage system should meet in this context. Second, we evaluated through performance analysis three "off-the-shelf" object store solutions, namely Rados, Cassandra and InterPlanetary File System (IPFS). In particular, we focused (i) on access times to push and get objects under different scenarios and (ii) on the amount of network traffic that is exchanged between the different sites during such operations. We also evaluated how the network latencies influence the access times and how the systems behave in case of network partitioning. Experiments have been conducted using the Yahoo Cloud System Benchmark (YCSB) on top of the Grid'5000 testbed. We showed that among the three tested solutions IPFS fills most of the criteria expected for a Fog/Edge computing infrastructure. [33], [32]

7.2.2. Renewable energy

With the emergence of the Future Internet and the dawning of new IT models such as cloud computing, the usage of data centers (DC), and consequently their power consumption, increase dramatically. Besides the ecological impact, the energy consumption is a predominant criterion for DC providers since it determines the daily cost of their infrastructure. As a consequence, power management becomes one of the main challenges for DC infrastructures and more generally for large-scale distributed systems. We have design the EpoCloud prototype, from hardware to middleware layers. This prototype aims at optimizing the energy consumption of mono-site Cloud DCs connected to the regular electrical grid and to renewable-energy sources. [17]

7.2.2.1. Green Energy awareness in SaaS Application.

With the proliferation of Cloud computing, data centers have to urgently face energy consumption issues. Although recent efforts such as the integration of renewable energy to data centers or energy efficient techniques in (virtual) machines contribute to the reduction of carbon footprint, creating green energy awareness around *Interactive Cloud Applications* by smartly using the presence of green energy has not been yet addressed. By *awareness*, we mean the inherited capability of SaaS applications to dynamically adapt with the availability of green energy and to reduce energy consumption while green energy is scarce or absent. In [25], we present two application controllers based on different metrics (e.g., availability of green energy, response time, user experience level). Based on extensive experiments with a real application benchmark and workloads in Grid'5000, results suggest that providers revenue can be increased as high as 64%, while 13% brown energy can be reduced without deprovisioning any physical or virtual resources at IaaS layer and 17 fold increment of performance can be guaranteed.

8. Bilateral Contracts and Grants with Industry

8.1. Cooperation with SIGMA group

Participants: Thomas Ledoux [correspondent], Simon Dupont.

In 2012, we have started a cooperation with Sigma Group (http://www.sigma.fr), a software editor and consulting enterprise. The cooperation consists in a joint (a so-called Cifre) PhD on eco-elasticity of software for the Cloud and the sponsorship of several engineering students at the MSc-level.

As a direct consequence of the increasing popularity of Cloud computing solutions, data centers are rapidly growing in number and size and have to urgently face with energy consumption issues. The aim of Simon Dupont's PhD, started in November 2012, is to explore the *software elasticity* capability in Software-as-a-Service (SaaS) development to promote the management of SaaS applications that are more flexible, more reactive to environment changes and therefore self-adaptive for a wider range of contexts. As a result, SaaS applications become more elastic and by transitivity more susceptible to energy constraints and optimization issues.

In 2016, Simon Dupont defended his PhD on "Cross-layer elasticity management for Cloud: towards an efficient usage of Cloud resources and services" [12]. Besides, we focused on ElaScript, a domain-specific language that offers Cloud administrators a simple and concise way to define complex elasticity-based reconfiguration plans [23].

9. Partnerships and Cooperations

9.1. Regional Initiatives

9.1.1. RFI Atlanstic 2020

9.1.1.1. CoMe4ACloud

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

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

CoMe4ACloud is an Atlanstic2020 funded project and supports a post-doc position. The project is led by Ascola research team and involves also AtlanModels and TASC, all of them from the LINA (Nantes Computer Science Laboratory) and situated at Ecole des Mines de Nantes. See https://come4acloud.github.io for more information.

9.1.2. Pays de la Loire

9.1.2.1. SyMeTRIC

Participant: Jean-Marc Menaud.

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

The long term goal of SyMeTRIC is to build a common Systems Medicine computing infrastructure to accelerate the discovery and validation of biomarkers in the fields of oncology, transplantation, and chronic cardiovascular diseases.

9.2. National Initiatives

9.2.1. CominLabs laboratory of excellence

9.2.1.1. EPOC

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

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

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

9.2.1.2. PrivGen

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

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

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

The partners are working on three main challenges:

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

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

9.2.1.3. SecCloud

Participants: Jacques Noyé [coordinator], Florent Marchand de Kerchove de Denterghem, Mario Südholt.

The high-level objective of the 3-year SecCloud (Secure Scripting for the Cloud) project is to enhance the security of devices on which web applications can be downloaded, i.e. to enhance client-side security in the context of the Cloud. In order to do so, the project relies on a language-based approach, focusing on three related issues:

- The definition of security policies for web architectures, especially on the client-side.
- Formally-proven analyses of web programming languages.
- Multi-level enforcement mechanisms for the security policies (based on static and dynamic analysis encompassing application-level and system-level software).

ASCOLA members are mainly interested in JavaScript as a programming language as well as the use of aspects as a seamless path from the definition of security policies and their composition to their implementation.

This year, we have finalized our proposal of extensible JavaScript modules and applied it to extend in a modular way the full-blown JavaScript interpreter Narcissus with several dynamic analyses including information-flow analyses.

9.2.2. ANR

9.2.2.1. SONGS (ANR/INFRA)

Participants: Adrien Lebre [coordinator], Jonathan Pastor, Anthony Simonet.

The SONGS project (Simulation of Next Generation Systems) is an ANR/INFRA project running for 48 months (starting in January 2012 with an allocated budget of 1.8MEuro, 95KEuro for ASCOLA).

The consortium is composed of 11 academic partners from Nancy (AlGorille, coordinator), Grenoble (MESCAL), Villeurbanne (IN2P3 Computing Center, GRAAL/Avalon - LIP), Bordeaux (CEPAGE, HiePACS, RUNTIME), Strasbourg (ICPS - LSIIT), Nantes (ASCOLA), Nice (MASCOTTE, MODALIS).

The goal of the SONGS project (http://infra-songs.gforge.inria.fr) is to extend the applicability of the SimGrid simulation framework from Grids and Peer-to-Peer systems to Clouds and High Performance Computation systems.

9.2.3. FSN

9.2.3.1. OpenCloudware (FSN)

Participants: Jean-Marc Menaud [coordinator], Thomas Ledoux.

The OpenCloudware project is coordinated by France Telecom, funded by the French Fonds National pour la Société Numérique (FSN, call Cloud n°1) and endorsed by competitiveness clusters Minalogic, Systematic and SCS. OpenCloudware is developed by a consortium of 18 partners bringing together industry and academic leaders, innovative technology start-ups and open source community expertise. The project started in 2012 for a duration of 42 months.

The OpenCloudware project aims at building an open software engineering platform, for the collaborative development of distributed applications to be deployed on multiple Cloud infrastructures. It will be available through a self-service portal. We target virtualized multi-tier applications such as JavaEE - OSGi. The results of OpenCloudware will contain a set of software components to manage the lifecycle of such applications, from modelling(Think), developing and building images (Build), to a multi-IaaS compliant PaaS platform (Run).

The ASCOLA project-team is mainly involved in the sub-projects "Think" (SLA model across Cloud layers) and "Run" (virtual machine manager for datacenters and placement constraints). The team has developed btrCloudStack, a private cloud based on the OpenSource CloudStack and integrating the work on placement rules and energy optimization. This software system has been extended this year.

9.2.3.2. Hosanna (FSN)

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

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

9.2.4. CPER

9.2.4.1. SeDuCe

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

The SeDuCe project (Sustainable Data Centers: Bring Sun, Wind and Cloud Back Together), aims to design an experimental infrastructure dedicated to the study of data centers with low energy footprint. This innovative data center will be the first experimental data center in the world for studying the energy impact of cloud computing and the contribution of renewable energy (solar panels, wind turbines) as well on the scientific, technological, that economical. This project is integrated in the national context of grid computing (Grid'5000), and the Constellation project, which will be an inter-node (Pays de la Loire, Brittany). He also participated in the validation of scientific work in interdisciplinary axis STIC and energy efficiency of the laboratory of excellence COMIN Labs.

9.2.5. Inria Project Labs

9.2.5.1. DISCOVERY

Participants: Ronan Alexandre Rcherreau, Adrien Lebre [coordinator], Anthony Simonet, Mario Südholt.

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

The DISCOVERY initiative [36] aims at exploring a new way of operating Utility Computing (UC) resources by leveraging any facilities available through the Internet in order to deliver widely distributed platforms that can better match the geographical dispersal of users as well as the ever increasing demand. Critical to the emergence of such locality-based UC (also referred as Fog/Edge Computing) platforms is the availability of appropriate operating mechanisms. The main objective of DISCOVERY is to design, implement, demonstrate and promote a new kind of Cloud Operting System (OS) that will enable the management of such a large-scale and widely distributed infrastructure in an unified and friendly manner.

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

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

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

9.2.6. InriaHub

9.2.6.1. MERCURY

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

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

9.3. European Initiatives

9.3.1. FP7 & H2020 Projects

9.3.1.1. CoqHoTT

Title: Coq for Homotopy Type Theory

Programm: H2020

Type: ERC

Duration: June 2015 - May 2020

Coordinator: Inria

Inria contact: Nicolas TABAREAU

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

9.3.1.2. BigStorage

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

Programm: H2020

Duration: January 2015 - December 2018

Coordinator: Universidad politecnica de Madrid

Partners:

Storage Research Group, Barcelona Supercomputing Center - Centro Nacional de Supercomputacion (Spain)

Ca Technologies Development Spain (Spain)

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

Deutsches Klimarechenzentrum (Germany)

ICS, Foundation for Research and Technology Hellas (Greece)

Fujitsu Technology Solutions (Germany)

Johannes Gutenberg Universitaet Mainz (Germany)

Universidad Politecnica de Madrid (Spain)

Seagate Systems Uk (United Kingdom)

Inria contact: G. Antoniu & A. Lebre

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

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events Organisation

10.1.1.1. Member of the Steering and Organizing Committees

- A. Lebre took part to the organisation of the Grid'5000 school in Grenoble (70 attendees).
- A. Lebre took part to the organisation of the workshop "Stockage informatique" during the Journées Scientifiques event in Nantes.
- J. Noyé was a co-organizer of LaMod '16, a workshop on Language Modularity co-located with Modularity '16 in Málaga, Spain.

10.1.2. Scientific Events Selection

10.1.2.1. Member of the Conference Program Committees

- J.-C. Royer was a member of the program committes WETCIE 2016, CAL 2016, ICIS 2016, IWAISE 2016, IIAI 2016.
- T. Ledoux was member of the program committees of the following workshops: Greens'16@ICSE, ARM'16@Middleware, CrossCloud'16@EuroSys.
- A. Lebre was a member of the program committees of ACM/IEEE CCGRID 2016, Europar 2016, ACM/IEEE SC 2016, IEEE CloudCom 2016, IEEE SSS 2016, OPTIM 2016.
- J. Noyé was a member of the program committee of Modularity '16 (Málaga, Spain).

10.1.3. Journal

10.1.3.1. Member of the Editorial Boards

- A. Lebre is associate editor for the IEEE Transactions on Big Data journal.
- M. Südholt is joint editor-in-chief of the journal Transactions on Modularity and Software Composition (Springer), formerly Transactions on AOSD.
- M. Südholt is an associate editor of the Journal on Programming, an open access journal.

10.1.3.2. Reviewer - Reviewing Activities

- A. Lebre has been a reviewer for the IEEE TPDS and IEEE TCC Journals, the IEEE Cloud Computing magazine, and the Journal of Parallel and Distributed Computing.
- T. Ledoux has been a reviewer for the IEEE Communications Letters.
- J. Noyé has been a reviewer for the Journal of Object Technology and Science of Computer Programming.

10.1.4. Invited Talks

- A. Lebre and Anthony Simonet have been invited to the 9th edition of the CloudControl Workshop (Sweden).
- A. Lebre has been invited to the BigStorage Initial Training Schoom (Spain).

10.1.5. Leadership within the Scientific Community

- A. Lebre is leading the OpenStack "Massively Distributed Working Group" (further information at: https://wiki.openstack.org/wiki/Massively_Distributed_Clouds).
- A. Lebre is member of the executive committee of the GDR CNRS RSD (Reseau et Système distribué). He is also co-leading the transversal action Virtualization and Clouds of this GDR.
- A. Lebre is member of the executive and architect committees of the Grid'5000 GIS (Groupement d'intérêt scientifique).
- T. Ledoux is member of the board of the Green Lab Center association. This association promotes and disseminates Green IT practices and research prototypes to the world of education, research and companies [0].
- M. Südholt is a member of the steering committees of the two international conferences on Programming and Modularity.
- M. Südholt has been a member of the board of the Aspect-Oriented Software Association.

10.1.6. Research Administration

- Pierre Cointe has been the head of the Lina laboratory that managed research in Computer Science of the main research institutions in Nantes.
- Jacques Noyé is the deputy head of the Mines Nantes department of Automation, Production and Computer Sciences.

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

The team is involved in the following undergraduate and graduate-level programs at Mines Nantes and University of Nantes (the institutions all of eaching staff belongs to):

- The team is a main contributor to the **engineering program of EMN**.
- Within this engineering program, the team is steering, chairing and the main contributor to a two-year **graduate-level informatics specialization**. H. Grall is managing this program.
- The team is leading a three-year **engineering program on software engineering**. T. Ledoux is managing this program.

The team has also been involved in the following MSc programs that have been carried out with partners from French and foreign universities:

- The team participates in the **MSc program "Alma"** on software architecture and distributed systems, a joint program steered by colleagues from University of Nantes. In this context, we are responsible for a 48-hour module on advanced software composition and take part in the program's governing board. M. Südholt is managing the participation of Mines Nantes in this program.
- J.-C. Royer was teaching "Architecture, component programming and OSGi", from March 7 until 12, level M1, at the University of Monastir (Tunisia).

m members have taught for about 220 hours on average in 2015 (hours of presence in front of students). Hereby, we have taken into account that researchers and some professors have not taught at times. In addition, another significant part of the program is taught by temporary staff, whose participation is managed by ASCOLA members.

[0] Green Lab Center

10.2.2. Juries

- J.-C. Royer was reviewer of the HDR of Mohamed Bhiri (Université Grenoble), September 7, 2016. He was also member of the PhD defense of Jonathan Pépin (Université de Nantes) December 5, Madhi Benmoussa (Université Paris XIII) December 6, and Amine Benelallam (Ecole des Mines) December 7.
- A. Lebre was a member of the PhD committee of Vincent Kherbache, "Ordonnancement des migrations a` chaud de machines virtuelles", Université de Nice - Sophia Antipolis, Dec. 7.
- T. Ledoux was a member of the PhD committee of Zakarea Al Shara (Univ. Montpellier), Nov. 17.

10.3. Popularization

- A. Lebre has been invited to the CargoDays event (Nantes).

11. Bibliography

Major publications by the team in recent years

[1] B. DE FRAINE, E. ERNST, M. SÜDHOLT.*Essential AOP: The A Calculus*, in "ACM Transactions on Programming Languages and Systems (TOPLAS)", December 2012, http://hal.inria.fr/hal-00676082.

[2] I. FIGUEROA, T. SCHRIJVERS, N. TABAREAU, É. TANTER.*Compositional Reasoning About Aspect Interference*, in "Modularity'14: 13th International Conference on Modularity", Lugano, Switzerland, April 2014, https://hal.inria.fr/hal-00919935.

[3] M. S. HASAN, Y. KOUKI, T. LEDOUX, J.-L. PAZAT.*Exploiting Renewable sources: when Green SLA becomes a possible reality in Cloud computing*, in "IEEE Transactions on Cloud Computing", July 2015, vol. PP, n^{o} 99, 1 [*DOI :* 10.1109/TCC.2015.2459710], https://hal.archives-ouvertes.fr/hal-01187907.

[4] G. JABER, G. LEWERTOWSKI, P.-M. PÉDROT, M. SOZEAU, N. TABAREAU.*The Definitional Side of the Forcing*, in "Logics in Computer Science", New York, United States, May 2016 [*DOI :* 10.1145/2933575.2935320], https://hal.archives-ouvertes.fr/hal-01319066.

[5] Y. KOUKI, F. ALVARES DE OLIVEIRA JR., S. DUPONT, T. LEDOUX.*A Language Support for Cloud Elasticity Management*, in "CCGrid'14: IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing", Chicago, United States, May 2014, p. 1-8, https://hal.archives-ouvertes.fr/hal-00941945.

[6] A. LEBRE, J. PASTOR, M. SÜDHOLT.*VMPlaceS: A Generic Tool to Investigate and Compare VM Placement Algorithms*, in "Europar 2015", Vienne, Austria, August 2015, https://hal.inria.fr/hal-01159033.

[7] R. POTTIER, J.-M. MENAUD.*TrustyDrive: a Multi-Cloud Storage Service that Protects Your Privacy*, in "IEEE 9th International Conference on Cloud Computing", San Francisco, United States, International Conference on Cloud Computing, June 2016, https://hal.inria.fr/hal-01322638.

[8] M. SOZEAU, N. TABAREAU.*Universe Polymorphism in Coq*, in "ITP'14: Interactive Theorem Proving", Vienna, Austria, July 2014, https://hal.inria.fr/hal-00974721.

[9] N. TABAREAU, M. SÜDHOLT, É. TANTER.*Aspectual Session Types*, in "Modularity'14 - 13th International Conference on Modularity", Lugano, Switzerland, April 2014, https://hal.inria.fr/hal-00872791.

[10] J. M. VAN HAM, G. SALVANESCHI, M. MEZINI, J. NOYÉ.*JEScala: Modular Coordination with Declarative Events and Joins*, in "Modularity' 14 - 13th International Conference on Modularity", Lugano, Switzerland, E. ERNST (editor), April 2014, https://hal.inria.fr/hal-00862332.

Publications of the year

Doctoral Dissertations and Habilitation Theses

[11] R.-A. CHERRUEAU.*A Compositional Language of Security Techniques for Information Privacy in the Cloud*, Ecole des Mines de Nantes, November 2016, https://tel.archives-ouvertes.fr/tel-01416166.

[12] S. DUPONT.*Crosslayer elasticity management for Cloud : towards an efficient usage of Cloud resources and services*, Ecole des Mines de Nantes, April 2016, https://tel.archives-ouvertes.fr/tel-01344377.

[13] F. MARCHAND DE KERCHOVE DE DENTERGHEM.*Extending interpreters by diverting, or how to extend interpreters without modifying their source code*, Ecole des Mines de Nantes, November 2016, https://tel.archives-ouvertes.fr/tel-01415588.

[14] J. PASTOR.*Contributions to massively distributed Cloud Computing infrastructures*, Ecole des Mines de Nantes, October 2016, https://tel.archives-ouvertes.fr/tel-01416099.

[15] N. TABAREAU.*Managing Logical and Computational Complexity using Program Transformations*, université de nantes, November 2016, Habilitation à diriger des recherches, https://tel.archives-ouvertes.fr/tel-01406351.

Articles in International Peer-Reviewed Journal

[16] N. BELDICEANU, M. CARLSSON, R. DOUENCE, H. SIMONIS.*Using finite transducers for describing and synthesising structural time-series constraints*, in "Constraints", January 2016, vol. 21, n^{o} 1 [*DOI :* 10.1007/s10601-015-9200-3], https://hal.inria.fr/hal-01370322.

[17] N. BELDICEANU, B. DUMAS FERIS, P. GRAVEY, S. HASAN, C. JARD, T. LEDOUX, Y. LI, D. LIME, G. MADI-WAMBA, J.-M. MENAUD, P. MOREL, M. MORVAN, M.-L. MOULINARD, A.-C. ORGERIE, J.-L. PAZAT, O. H. ROUX, A. SHARAIHA.*Towards energy-proportional Clouds partially powered by renewable energy*, in "Computing", January 2017, vol. 99, n^{o} 1, 20 [*DOI :* 10.1007/s00607-016-0503-z], https://hal.inria.fr/hal-01340318.

[18] I. FIGUEROA, N. TABAREAU, É. TANTER.*Effect capabilities for Haskell: Taming effect interference in monadic programming*, in "Science of Computer Programming", April 2016, vol. 119, p. 3-30 [*DOI :* 10.1016/j.scico.2015.11.010], https://hal.inria.fr/hal-01400002.

[19] D. SERRANO, S. BOUCHENAK, Y. KOUKI, F. ALVARES DE OLIVEIRA JR., T. LEDOUX, J. LEJEUNE, J. SOPENA, L. ARANTES, P. SENS.*SLA guarantees for cloud services*, in "Future Generation Computer Systems", January 2016, vol. 54, p. 233–246 [*DOI :* 10.1016/j.future.2015.03.018], https://hal.archives-ouvertes.fr/hal-01162654.

International Conferences with Proceedings

[20] E. ARAFAILOVA, N. BELDICEANU, R. DOUENCE, P. FLENER, M. A. FRANCISCO RODRÍGUEZ, J. PEARSON, H. SIMONIS.*Time-Series Constraints: Improvements and Application in CP and MIP Contexts*, in "CPAIOR 2016 - 13th International Conference on Integration of Artificial Intelligence and Operations

Research Techniques in Constraint Programming", Banff, Canada, C.-G. QUIMPER (editor), Lecture Notes in Computer Science, Springer, May 2016, vol. 9676, p. 18-34 [*DOI :* 10.1007/978-3-319-33954-2], https://hal.inria.fr/hal-01355262.

[21] J. COHEN.*Renaming Global Variables in C Mechanically Proved Correct*, in "Fourth International Workshop on Verification and Program Transformation", Eindhoven, Netherlands, April 2016, https://hal.archives-ouvertes.fr/hal-01277269.

[22] P.-E. DAGAND, N. TABAREAU, É. TANTER.*Partial Type Equivalences for Verified Dependent Interoperability*, in "ICFP 2016 - 21st ACM SIGPLAN International Conference on Functional Programming", Nara, Japan, September 2016, p. 298-310 [*DOI :* 10.1145/2951913.2951933], https://hal.inria.fr/hal-01328012.

[23] S. DUPONT, S. BOURI, F. ALVARES DE OLIVEIRA, T. LEDOUX.*ElaScript: a DSL for Coding Elasticity in Cloud Computing*, in "32nd ACM Symposium on Applied Computing - Track on Cloud Computing", Marrakesh, Morocco, Proceedings of the 32nd ACM Symposium on Applied Computing - Track on Cloud Computing, April 2017, https://hal.archives-ouvertes.fr/hal-01400236.

[24] A. GARNIER, J.-M. MENAUD, N. MONTAVONT.*Bringing Complex Event Processing into Multitree Modelling of Sensors*, in "International Conference on Distributed Applications and Interoperable Systems (DAIS)", Heraklion, Greece, June 2016, https://hal.inria.fr/hal-01322670.

[25] M. S. HASAN, F. ALVARES DE OLIVEIRA, T. LEDOUX, J.-L. PAZAT.*Enabling Green Energy awareness in Interactive Cloud Application*, in "IEEE International Conference on Cloud Computing Technology and Science 2016", Luxembourg, Luxembourg, December 2016, https://hal.inria.fr/hal-01365230.

[26] G. JABER, G. LEWERTOWSKI, P.-M. PÉDROT, M. SOZEAU, N. TABAREAU.*The Definitional Side of the Forcing*, in "Logics in Computer Science", New York, United States, May 2016 [*DOI :* 10.1145/2933575.2935320], https://hal.archives-ouvertes.fr/hal-01319066.

[27] A. LEBRE, J. PASTOR, A. SIMONET, F. DESPREZ.*Revising OpenStack to Operate Fog/Edge Computing infrastructures*, in "IEEE International Conference on Cloud Engineering", Vancouver, France, April 2017, https://hal.inria.fr/hal-01273427.

[28] Y. LI, A.-C. ORGERIE, J.-M. MENAUD.*Balancing the use of batteries and opportunistic scheduling policies for maximizing renewable energy consumption in a Cloud data center*, in "PDP 2017 - 25th Euromicro International Conference on Parallel, Distributed, and Network-Based Processing", St Petersburg, Russia, March 2017, https://hal.inria.fr/hal-01432752.

[29] A. LÈBRE, A. SIMONET, A.-C. ORGERIE.*Deploying Distributed Cloud Infrastructures: Who and at What Cost?*, in "Proceedings of the fifth IEEE International Workshop on Cloud Computing Interclouds, Multiclouds, Federations, and Interoperability", Berlin, Germany, Intercloud 2016, April 2016, 6 [*DOI :* 10.1109/IC2EW.2016.48], https://hal.inria.fr/hal-01404594.

[30] R. POTTIER, J.-M. MENAUD.*TrustyDrive: a Multi-Cloud Storage Service that Protects Your Privacy*, in "IEEE 9th International Conference on Cloud Computing", San Francisco, United States, International Conference on Cloud Computing, June 2016, https://hal.inria.fr/hal-01322638.

[31] J.-C. ROYER, A. SANTANA DE OLIVEIRA.*AAL and Static Conflict Detection in Policy*, in "15th International Conference on Cryptology and Network Security", Milan, Italy, S. FORESTI, G. PERSIANO (editors),

Cryptology and Network Security, Springer, November 2016, n^o 10052, p. 362-382 [*DOI :* 10.1007/978-3-319-48965-0_22], https://hal.archives-ouvertes.fr/hal-01396376.

Conferences without Proceedings

[32] B. CONFAIS, A. LÈBRE, B. PARREIN.*Performance Analysis of Object Store Systems in a Fog/Edge Computing Infrastructures*, in "CloudCom", Luxembourg, Luxembourg, December 2016, https://hal.archives-ouvertes.fr/hal-01397686.

[33] B. CONFAIS, A. LÈBRE, B. PARREIN.*Quel système de stockage pour les architectures Fog ?*, in "Compas'2016", Lorient, France, July 2016, https://hal.archives-ouvertes.fr/hal-01376292.

Scientific Books (or Scientific Book chapters)

[34] Y. KOUKI, F. ALVARES DE OLIVEIRA, T. LEDOUX.*Cloud Capacity Planning and Management*, in "Encyclopedia of Cloud Computing", Wiley-IEEE Press, July 2016, https://hal.archives-ouvertes.fr/hal-01342153.

Research Reports

[35] R.-A. CHERRUEAU, A. LEBRE, D. PERTIN, A. SIMONET, M. SIMONIN.*ENOS: a HolisticFramework forConducting ScientificEvaluations of OpenStack*, Inria Rennes Bretagne Atlantique ; Nantes, November 2016, n^o RT-0485, https://hal.inria.fr/hal-01415522.

[36] A. LÈBRE, J. PASTOR, F. DESPREZ.*A Ring to Rule Them All - Revising OpenStack Internals to Operate Massively Distributed Clouds: The Discovery Initiative - Where Do We Are ?*, Inria, February 2016, n^o RT-0480, p. 1-24, https://hal.inria.fr/hal-01320235.

Other Publications

[37] B. AHRENS, R. MATTHES, A. MÖRTBERG.*From signatures to monads in UniMath*, December 2016, working paper or preprint, https://hal.inria.fr/hal-01410487.

[38] W. BENGHABRIT, H. GRALL, J.-C. ROYER.*Monitoring accountability policies with AccMon framework*, June 2016, GDR-GPL, Poster, https://hal.inria.fr/hal-01332040.

[39] F. MARCHAND DE KERCHOVE, J. NOYÉ, M. SÜDHOLT.*Extensible Modules for JavaScript*, April 2016, 3, SAC '16 - 31st Annual ACM Symposium on Applied Computing, Poster [*DOI :* 10.1145/2851613.2851958], https://hal.inria.fr/hal-01407340.

[40] P.-M. PÉDROT, N. TABAREAU.*An Effectful Way to Eliminate Addiction to Dependence*, January 2017, working paper or preprint, https://hal.inria.fr/hal-01441829.

References in notes

[41] M. AKŞIT, S. CLARKE, T. ELRAD, R. E. FILMAN (editors). *Aspect-Oriented Software Development*, Addison-Wesley Professional, September 2004.

[42] C. ALLAN, P. AVGUSTINOV, A. S. CHRISTENSEN, L. HENDREN, S. KUZINS, O. LHOTÁK, O. DE MOOR, D. SERENI, G. SITTAMPALAM, J. TIBBLE.*Adding trace matching with free variables to AspectJ*, in "ACM

Conference on Object-Oriented Programming, Systems and Languages (OOPSLA)", R. P. GABRIEL (editor), ACM Press, 2005.

[43] R. ALLEN, D. GARLAN.*A Formal Basis for Architectural Connection*, in "ACM Transactions on Software Engineering and Methodology", July 1997, vol. 6, n° 3, p. 213–49.

[44] J. H. ANDREWS.*Process-Algebraic Foundations of Aspect-Oriented Programming*, in "Proceedings of the 3rd International Conference on Metalevel Architectures and Separation of Crosscutting Concerns", Lecture Notes in Computer Science, 2001, vol. 2192, p. 187–209.

[45] M. ARMBRUST, A. FOX, R. GRIFFITH, A. D. JOSEPH, R. KATZ, A. KONWINSKI, G. LEE, D. PATTERSON, A. RABKIN, I. STOICA, M. ZAHARIA.*A view of cloud computing*, in "Communications of the ACM", 2010, vol. 53, n° 4, p. 50–58.

[46] T. H. AUSTIN, C. FLANAGAN.*Multiple facets for dynamic information flow*, in "Proceedings of the 39th annual ACM SIGPLAN-SIGACT symposium on Principles of programming languages", New York, USA, POPL '12, ACM, 2012, p. 165–178, http://doi.acm.org/10.1145/2103656.2103677.

[47] A. BELOGLAZOV, R. BUYYA.*Energy efficient resource management in virtualized cloud data centers*, in "in: Proceedings of the 2010 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing, CCGRID'10, IEEE Computer Society", 2010, p. 826–831.

[48] L. D. BENAVIDES NAVARRO, M. SÜDHOLT, W. VANDERPERREN, B. DE FRAINE, D. SUVÉE.*Explicitly distributed AOP using AWED*, in "Aspect-Oriented Software Development (AOSD)", ACM Press, March 2006, p. 51-62.

[49] G. S. BLAIR, G. COULSON, P. ROBIN, M. PAPATHOMAS.*An architecture for next generation middleware*, in "Proceedings of the IFIP International Conference on Distributed Systems Platforms and Open Distributed Processing", Springer-Verlag, 1998.

[50] A. BRACCIALIA, A. BROGI, C. CANAL.*A formal approach to component adaptation*, in "Journal of Systems and Software", 2005.

[51] R. E. BROWN, E. R. MASANET, B. NORDMAN, W. F. TSCHUDI, A. SHEHABI, J. STANLEY, J. G. KOOMEY, D. A. SARTOR, P. T. CHAN.*Server and Data Center Energy Efficiency: Public Law 109-431*, in "Report to Congress", 06/2008 2008.

[52] S. CAPECCHI, I. CASTELLANI, M. DEZANI-CIANCAGLINI, T. REZK.*Session Types for Access and Information Flow Control*, in "CONCUR 2010 - Concurrency Theory, 21th International Conference, CONCUR 2010, Paris, France, August 31-September 3, 2010. Proceedings", P. GASTIN, F. LAROUSSINIE (editors), Lecture Notes in Computer Science, Springer, 2010, vol. 6269, p. 237–252, http://dx.doi.org/10.1007/978-3-642-15375-4_17.

[53] E. M. CLARKE, O. GRUMBERG, D. A. PELED.*Model Checking*, The MIT Press, Cambridge, Massachusetts, 1999.

[54] A. COLYER, A. CLEMENT.*Large-scale AOSD for Middleware*, in "Proceedings of the 3rd ACM Int. Conf. on Aspect-Oriented Software Development (AOSD), Lancaster", K. LIEBERHERR (editor), ACM Press, 2004, p. 56–65.

[55] F. DEREMER, H. H. KRON.*Programming-in-the-large versus programming-in-the-small*, in "IEEE Transactions on Software Engineering", 1976, vol. SE-2, n^{o} 2, p. 80-86.

[56] G. DECKER, O. KOPP, F. LEYMANN, M. WESKE.*BPEL4Chor: Extending BPEL for Modeling Choreographies*, in "IEEE International Conference on Web Services (ICWS 2007)", IEEE Computer Society, 2007, p. 296–303.

[57] E. W. DIJKSTRA.*On the role of scientific thought*, in "Selected Writings on Computing: A Personal Perspective", Springer-Verlag, 1974, p. 60–66, Published in 1982.

[58] R. DOUENCE, P. FRADET, M. SÜDHOLT.*A framework for the detection and resolution of aspect interactions*, in "Proceedings of the ACM SIGPLAN/SIGSOFT Conference on Generative Programming and Component Engineering (GPCE'02)", Lecture Notes in Computer Science, Springer-Verlag, October 2002, vol. 2487, p. 173–188, http://hal.inria.fr/inria-00072153.

[59] R. DOUENCE, P. FRADET, M. SÜDHOLT.*Trace-Based Aspects*, in "Aspect-Oriented Software Development", M. AKŞIT, S. CLARKE, T. ELRAD, R. E. FILMAN (editors), Addison-Wesley, 2004, p. 201-218.

[60] R. DOUENCE, O. MOTELET, M. SÜDHOLT.*A formal definition of crosscuts*, in "Proceedings of the 3rd International Conference on Metalevel Architectures and Separation of Crosscutting Concerns", Lecture Notes in Computer Science, Springer-Verlag, 2001, vol. 2192, p. 170–186.

[61] R. DOUENCE, D. LE BOTLAN, J. NOYÉ, M. SÜDHOLT.*Concurrent Aspects*, in "Proc. of the Int. ACM Conf. on Generative Programming and Component Engineering (GPCE)", ACM Press, October 2006, p. 79-88.

[62] H. FOSTER, S. UCHITEL, J. MAGEE, J. KRAMER.*Model-based Verification of Web Service Compositions*, in "Proceedings of the 18th IEEE Int. Conf. on Automated Software Engineering (ASE'03)", IEEE Computer Society, 2003, p. 152–163.

[63] I. FOSTER, Y. ZHAO, I. RAICU, S. LU.*Cloud computing and grid computing 360-degree compared*, in "Grid Computing Environments Workshop, 2008. GCE'08", Ieee, 2008, p. 1–10.

[64] A. FUGGETTA, G. P. PICCO, G. VIGNA.*Understanding Code Mobility*, in "IEEE Transactions on Software Engineering", May 1998, vol. 24, n^{o} 5, p. 342–361.

[65] GREENPEACE.*Make IT green: Cloud computing and its contribution to climate change*, Greenpeace International, March 2010.

[66] K. HONDA, N. YOSHIDA, M. CARBONE.*Multiparty asynchronous session types*, in "Proceedings of the 35th ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages, POPL 2008, San Francisco, California, USA, January 7-12, 2008", G. C. NECULA, P. WADLER (editors), ACM, 2008, p. 273–284, http://www.doc.ic.ac.uk/~yoshida/multiparty/multiparty.pdf.

[67] G. KICZALES, E. HILSDALE, J. HUGUNIN, M. KERSTEN, J. PALM, W. G. GRISWOLD.*An Overview of AspectJ*, in "ECOOP 2001 — Object-Oriented Programming 15th European Conference, Budapest Hungary", Berlin, J. L. KNUDSEN (editor), Lecture Notes in Computer Science, Springer-Verlag, June 2001, vol. 2072, p. 327–353, http://www.eclipse.org/aspectj/.

[68] G. KICZALES.*Aspect Oriented Programming*, in "Proc. of the Int. Workshop on Composability Issues in Object-Orientation (CIOO'96) at ECOOP", July 1996, Selected paper published by dpunkt press, Heidelberg, Germany.

[69] G. KICZALES, J. DES RIVIERES, DANIEL G. BOBROW.*The Art of the Meta-Object Protocol*, MIT Press, Cambridge (MA), USA, 1991.

[70] J. KIENZLE, R. GUERRAOUI.*AOP - Does It Make Sense? The Case of Concurrency and Failures*, in "16th European Conference on Object-Oriented Programming (ECOOP'2002)", Malaga, Spain, B. MAGNUSSON (editor), Lecture Notes in Computer Science, Springer-Verlag, 2002.

[71] T. LEDOUX.*OpenCorba: a Reflective Open Broker*, in "ACM Meta-Level Architectures and Reflection, Second International Conference, Reflection'99", Saint-Malo, France, P. COINTE (editor), Lecture Notes in Computer Science, Springer-Verlag, July 1999, vol. 1616, p. 197–214.

[72] M. MCILROY.*Mass produced software components*, in "Mass produced software components", Garmish, Germany, P. NAUR, B. RANDELL (editors), NATO Science Committee, October 1968, p. 138-155.

[73] N. MEDVIDOVIC, R. N. TAYLOR.*A Classification and Comparison Framework for Software Architecture Description Languages*, in "IEEE Transactions on Software Engineering", January 2000, vol. 26, n^{o} 1, p. 70-93.

[74] M. MERNIK, J. HEERING, A. M. SLOANE.*When and How to Develop Domain-Specific Languages*, in "ACM Computing Surveys", December 2005, vol. 37, n^{o} 4, p. 316-344.

[75] L. MIKHAJLOV, E. SEKERINSKI.*A study of the fragile base class*, in "A study of the fragile base class", Brussels, Belgium, E. JUL (editor), Lecture Notes in Computer Science, July 1998, vol. 1445, p. 355-382.

[76] D. H. NGUYEN, M. SÜDHOLT.*VPA-based aspects: better support for AOP over protocols*, in "4th IEEE International Conference on Software Engineering and Formal Methods (SEFM'06)", IEEE Computer Society Press, September 2006.

[77] O. NIERSTRASZ.*Regular Types for Active Objects*, in "Object-Oriented Software Composition", O. NIERSTRASZ, D. TSICHRITZIS (editors), Prentice Hall, 1995, chap. 4, p. 99–121.

[78] M. NISHIZAWA, S. CHIBA, M. TATSUBORI.*Remote Pointcut - A Language Construct for Distributed AOP*, in "Proceedings of the 3rd ACM Int. Conf. on Aspect-Oriented Software Development (AOSD), Lancaster", ACM Press, 2004.

[79] D. L. PARNAS.*On the criteria for decomposing systems into modules*, in "Communications of the ACM", December 1972, vol. 15, n^{o} 12, p. 1053-1058.

[80] S. PEARSON.*Toward Accountability in the Cloud*, in "Internet Computing, IEEE", July-Aug. 2011, vol. 15, n^{o} 4, p. 64-69, http://dx.doi.org/10.1109/MIC.2011.98.

[81] F. PLASIL, S. VISNOVSKY.*Behavior Protocols for Software Components*, in "Transactions on Software Engineering", January 2002, vol. 28, n^{o} 9.

[82] F. PUNTIGAM.*Coordination Requirements Expressed in Types for Active Objects*, in "ECOOP'97—Object-Oriented Programming", M. AKŞIT, S. MATSUOKA (editors), Lecture Notes in Computer Science, Springer-Verlag, 1997, vol. 1241, p. 367–388.

[83] N. SHARMA, S. BARKER, D. IRWIN, P. SHENOY.*Blink: managing server clusters on intermittent power*, in "SIGARCH Comput. Archit. News", March 2011, vol. 39, p. 185–198, http://dx.doi.org/10.1145/1961295.1950389.

[84] M. SHAW, D. GARLAN.*Software Architecture: Perspectives on an Emerging Discipline*, Prentice-Hall, 1996.

[85] B. C. SMITH.*Reflection and Semantics in LISP*, Xerox Palto Alto Research Center, Palo Alto, 1984, n^o P84-00030.

[86] S. SOARES, E. LAUREANO, P. BORBA.*Implementing distribution and persistence aspects with AspectJ* , in "Proceedings of the 17th ACM conference on Object-oriented programming, systems, languages, and applications (OOPSLA-02)", C. NORRIS, J. J. B. FENWICK (editors), ACM SIGPLAN Notices, ACM Press, November 4–8 2002, vol. 37, 11, p. 174–190.

[87] S. SUNDARESWARAN.*Ensuring Distributed Accountability for Data Sharing in the Cloud*, in "Dependable and Secure Computing", 2012, vol. 9, http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6165313.

[88] R. J. WALKER, K. VIGGERS.*Implementing Protocols via Declarative Event Patterns*, in "Proceedings of the ACM SIGSOFT International Symposium on Foundations of Software Engineering (FSE-12)", ACM Press, 2004, p. 159 - 169.

[89] M. WAND, G. KICZALES, C. DUTCHYN.*A Semantics for Advice and Dynamic Join Points in Aspect-Oriented Programming*, in "ACM Transactions on Programming Languages and Systems (TOPLAS)", 2004, vol. 26, n^o 5, p. 890–910.

[90] D. M. YELLIN, R. E. STROM.*Protocol specifications and component adaptors*, in "ACM Transactions of Programming Languages and Systems", March 1997, vol. 19, n^o 2, p. 292–333.

[91] Q. ZHANG, L. CHENG, R. BOUTABA.*Cloud computing: state-of-the-art and research challenges*, in "Journal of internet services and applications", 2010, vol. 1, n^o 1, p. 7–18.

[92] A. VAN DEURSEN, P. KLINT, J. VISSER.*Domain-Specific Languages: An Annotated Bibliography*, in "ACM SIGPLAN Notices", June 2000, vol. 35, n^o 6, p. 26-36.

Project-Team ASPI

Applications of interacting particle systems to statistics

IN COLLABORATION WITH: Institut de recherche mathématique de Rennes (IRMAR)

IN PARTNERSHIP WITH:
CNRS

Université Haute Bretagne (Rennes 2)

Université Rennes 1

RESEARCH CENTER
Rennes - Bretagne-Atlantique

THEME
Stochastic approaches

Table of contents

Project-Team ASPI

Creation of the Project-Team: 2005 January 10, updated into Team: 2017 January 01

Keywords:

Computer Science and Digital Science:

3.4.5. - Bayesian methods
3.4.7. - Kernel methods
5.4.5. - Object tracking and motion analysis
5.4.6. - Object localization
5.9.2. - Estimation, modeling
6.1.1. - Continuous Modeling (PDE, ODE)
6.1.2. - Stochastic Modeling (SPDE, SDE)
6.2.2. - Numerical probability
6.2.3. - Probabilistic methods
6.2.4. - Statistical methods
6.2.5. - Numerical Linear Algebra
6.3.2. - Data assimilation
6.3.4. - Model reduction
6.3.5. - Uncertainty Quantification

Other Research Topics and Application Domains:

2.3. - Epidemiology
3.2. - Climate and meteorology
3.3.2. - Water: sea & ocean, lake & river
3.3.4. - Atmosphere
9.4.3. - Physics
9.4.4. - Chemistry
9.9. - Risk management

1. Members

Research Scientists

Frédéric Cérou [Inria, Researcher]
Patrick Héas [Inria, Researcher]
François Le Gland [Team leader, Inria, Senior Researcher]
Mathias Rousset [Inria, Researcher, since Sep 2016, HDR]

Faculty Member

Valérie Monbet [Univ. Rennes 1, Professor, HDR]

PhD Students

Thi Tuyet Trang Chau [Univ. Rennes 1]
Audrey Cuillery [DCNS Nantes, CIFRE grant, since Apr 2016]
Ramatoulaye Dabo [Univ. Assane Seck de Ziguinchor and Univ. Rennes 1, co–tutelle]
Marie Morvan [Univ. Rennes 1]
Audrey Poterie [Univ. Rennes 1]
Kersane Zoubert–Ousseni [CEA Grenoble]

Post-Doctoral Fellow
Hassan Maatouk [Inria, since Sep 2016]

Administrative Assistant
Fabienne Cuyollaa [Inria]

Other
Arnaud Guyader [Univ. Paris VI, Professor, HDR]

2. Overall Objectives

2.1. Overall Objectives

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

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

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

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

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

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

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

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

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

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

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

3. Research Program

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

Monte Carlo methods are numerical methods that are widely used in situations where (i) a stochastic (usually Markovian) model is given for some underlying process, and (ii) some quantity of interest should be evaluated, that can be expressed in terms of the expected value of a functional of the process trajectory, which includes as an important special case the probability that a given event has occurred. Numerous examples can be found, e.g. in financial engineering (pricing of options and derivative securities) [43], in performance evaluation of communication networks (probability of buffer overflow), in statistics of hidden Markov models (state estimation, evaluation of contrast and score functions), etc. Very often in practice, no analytical expression is available for the quantity of interest, but it is possible to simulate trajectories of the underlying process. The idea behind Monte Carlo methods is to generate independent trajectories of this process or of an alternate instrumental process, and to build an approximation (estimator) of the quantity of interest in terms of the weighted empirical probability distribution associated with the resulting independent sample. By the law of large numbers, the above estimator converges as the size N of the sample goes to infinity, with rate $1/\sqrt{N}$ and the asymptotic variance can be estimated using an appropriate central limit theorem. To reduce the variance of the estimator, many variance reduction techniques have been proposed. Still, running independent Monte Carlo simulations can lead to very poor results, because trajectories are generated *blindly*, and only afterwards are the corresponding weights evaluated. Some of the weights can happen to be negligible, in which case the corresponding trajectories are not going to contribute to the estimator, i.e. computing power has been wasted.

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

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

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

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

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

In the special case of particle filtering, originally developed within the tracking community, the algorithms yield a numerical approximation of the optimal Bayesian filter, i.e. of the conditional probability distribution of the hidden state given the past observations, as a (possibly weighted) empirical probability distribution of the system of particles. In its simplest version, introduced in several different scientific communities under the name of *bootstrap filter* [45], *Monte Carlo filter* [49] or *condensation* (conditional density propagation) algorithm [48], and which historically has been the first algorithm to include a resampling step, the selection mechanism is governed by the likelihood function: at each time step, a particle is more likely to survive and to replicate at the next generation if it is consistent with the current observation. The algorithms also provide as a by–product a numerical approximation of the likelihood function, and of many other contrast functions for parameter estimation in hidden Markov models, such as the prediction error or the conditional least–squares criterion.

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

> positioning, navigation, and tracking [46], [39], visual tracking [48], mobile robotics [40], [61], ubiquitous computing and ambient intelligence, sensor networks, risk evaluation and simulation of rare events [44], genetics, molecular simulation [41], etc.

Other examples of the many applications of particle filtering can be found in the contributed volume [28] and in the special issue of *IEEE Transactions on Signal Processing* devoted to *Monte Carlo Methods for Statistical Signal Processing* in February 2002, where the tutorial paper [29] can be found, and in the textbook [56] devoted to applications in target tracking. Applications of sequential Monte Carlo methods to other areas, beyond signal and image processing, e.g. to genetics, can be found in [54]. A recent overview can also be found in [31].

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

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

These algorithms can be interpreted as numerical approximation schemes for Feynman–Kac distributions, a pathwise generalization of Gibbs–Boltzmann distributions, in terms of the weighted empirical probability distribution associated with a system of particles. This abstract point of view [36], [35], has proved to be extremely fruitful in providing a very general framework to the design and analysis of numerical approximation schemes, based on systems of branching and / or interacting particles, for nonlinear dynamical systems with values in the space of probability distributions, associated with Feynman–Kac distributions. Many asymptotic results have been proved as the number N of particles (sample size) goes to infinity, using techniques coming from applied probability (interacting particle systems, empirical processes [63]), see e.g. the survey article [36] or the textbooks [35], [34], and references therein

> convergence in $\mathbb{L}^p$, convergence as empirical processes indexed by classes of functions, uniform convergence in time, see also [52], [53], central limit theorem, see also [50], [37], propagation of chaos, large deviations principle, etc.

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

3.2. Multilevel splitting for rare event simulation

See 4.2, and 6.1.

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

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

In such complex systems, analytical methods cannot be used, and naive Monte Carlo methods are clearly unefficient to estimate accurately very small probabilities. Besides importance sampling, an alternate widespread technique consists in multilevel splitting [51], where trajectories going towards the critical set are given offsprings, thus increasing the number of trajectories that eventually reach the critical set. As shown in [6], the Feynman–Kac formalism of 3.1 is well suited for the design and analysis of splitting algorithms for rare event simulation.

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

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

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

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

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

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

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

Just as in the static case, the algorithm learns

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

Contributions have been given to

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

and in the static case, to

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

A related issue is global optimization. Indeed, the difficult problem of finding the set M of global minima of a real–valued function V can be replaced by the apparently simpler problem of sampling a population from a probability distribution depending on a small parameter, and asymptotically supported by the set M as the small parameter goes to zero. The usual approach here is to use the cross–entropy method [57], [33], which relies on learning the optimal importance distribution within a prescribed parametric family. On the other hand, multilevel splitting methods could provide an alternate nonparametric approach to this problem.

3.3. Statistical learning: pattern recognition and nonparametric regression

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

In general, there is no way to guess exactly the value of the feature associated with the new object, and the minimal error that can be done is that of the Bayes estimator, which cannot be computed by lack of knowledge of the distribution of the object–feature pair, but the Bayes estimator can be useful to characterize the strength of the method. So the best that can be expected is that the NN estimator converges, say when the sample size N grows, to the Bayes estimator. This is what has been proved in great generality by Stone [58] for the mean square convergence, provided that the object is a finite–dimensional random variable, the feature is a square–integrable random variable, and the ratio k/N goes to 0. Nearest neighbor estimator is not the only local averaging estimator with this property, but it is arguably the simplest.

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

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

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

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

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

that would be a source of interesting problems.

4. Application Domains

4.1. Localisation, navigation and tracking

See 7.1.

Among the many application domains of particle methods, or interacting Monte Carlo methods, ASPI has decided to focus on applications in localisation (or positioning), navigation and tracking [46], [39], which already covers a very broad spectrum of application domains. The objective here is to estimate the position (and also velocity, attitude, etc.) of a mobile object, from the combination of different sources of information, including

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

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

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

which particle methods can easily handle (map-matching). This Bayesian dynamical estimation problem is also called filtering, and its numerical implementation using particle methods, known as particle filtering, has been introduced by the target tracking community [45], [56], which has already contributed to many of the most interesting algorithmic improvements and is still very active, and has found applications in

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

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

4.2. Rare event simulation

See 3.2, and 6.1.

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

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

In such complex systems, analytical methods cannot be used, and naive Monte Carlo methods are clearly unefficient to estimate accurately very small probabilities. Besides importance sampling, an alternate widespread technique consists in multilevel splitting [51], where trajectories going towards the critical set are given offsprings, thus increasing the number of trajectories that eventually reach the critical set. This approach not only makes it possible to estimate the probability of the rare event, but also provides realizations of the random trajectory, given that it reaches the critical set, i.e. provides realizations of typical critical trajectories, an important feature that methods based on importance sampling usually miss.

ASPI is contributing (or has contributed recently) to several applications of multilevel splitting for rare event simulation, such as risk assessment in air traffic management, detection in sensor networks, and protection of digital documents.

5. Highlights of the Year

5.1. Highlights of the Year

Frédéric Cérou and Arnaud Guyader have received the prize of the best recent paper published in the journal *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques* for their joint paper [3] in collaboration with Gérard Biau (université Pierre et Marie Curie). This paper analyzes ABC (approximate Bayesian computation) — a family of computational techniques which offer an almost automated solution in situations where evaluation of the likelihood is computationally prohibitive, or whenever suitable likelihoods are not available — from the point of view of k—nearest neighbor theory and it explores the statistical properties of its outputs. The paper discusses in particular some asymptotic features of the genuine conditional density estimate associated with ABC, which is an interesting hybrid between a k–nearest neighbor and a kernel method.

6. New Results

6.1. Central limit theorem for adaptive multilevel splitting

Participants: Frédéric Cérou, Arnaud Guyader, Mathias Rousset.

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

In this work, we consider the adaptive multilevel splitting algorithm as a Fleming–Viot particle system: the particles are indexed by levels instead of time, and the associated states are given by first entrance into level sets, in a similar fashion as in [38]. A rigorous proof of a central limit theorem has been obtained in [24] for Fleming–Viot particle systems. The application to AMS (adaptive multilevel splitting) algorithm is in preparation.

6.2. An efficient algorithm for video super-resolution based on a sequential model

Participant: Patrick Héas.

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

In [16], we propose a novel procedure for video super–resolution, that is the recovery of a sequence of high–resolution images from its low-resolution counterpart. Our approach is based on a "sequential" model (i.e., each high-resolution frame is supposed to be a displaced version of the preceding one) and considers the use of sparsity-enforcing priors. Both the recovery of the high–resolution images and the motion fields relating them is tackled. This leads to a large–dimensional, non-convex and non-smooth problem. We propose an algorithmic framework to address the latter. Our approach relies on fast gradient evaluation methods and modern optimization techniques for non-differentiable/non-convex problems. Unlike some other previous works, we show that there exists a provably-convergent method with a complexity linear in the problem dimensions. We assess the proposed optimization method on several video benchmarks and emphasize its good performance with respect to the state of the art.

6.3. Low–rank approximation and dynamic mode decomposition

Participant: Patrick Héas.

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

Dynamic mode decomposition (DMD) has emerged as a powerful tool for analyzing the dynamics of non-linear systems from experimental datasets. Recently, several attempts have extended DMD to the context of low-rank approximations. This low-rank extension takes the form of a non–convex optimization problem. To the best of our knowledge, only sub–optimal algorithms have been proposed in the literature to compute the solution of this problem. In [26], we prove that there exists a closed-form optimal solution to this problem and design an effective algorithm to compute it based on singular value decomposition (SVD). Based on this solution, we then propose efficient procedures for reduced-order modeling and for the identification of the low-rank DMD modes and amplitudes. Experiments illustrates the gain in performance of the proposed algorithm compared to state-of-the-art techniques.

6.4. Model reduction from partial observations

Participant: Patrick Héas.

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

In [25], we deal with model order reduction of parametric partial differential equations (PPDE). We consider the specific setup where the solutions of the PPDE are only observed through a partial observation operator and address the task of finding a good approximation subspace of the solution manifold. We provide and study several tools to tackle this problem. We first identify the best worst–case performance achievable in this setup and propose simple procedures to approximate this optimal solution. We then provide, in a simplified setup, a theoretical analysis relating the achievable reduction performance to the choice of the observation operator and the prior knowledge available on the solution manifold.

In [22], we focus on reduced modeling of dynamical systems, in an analogous partial observation setup. Assuming prior knowledge available, we provide a unified reduction framework based on an a posteriori characterisation of the uncertainties on the solution manifold. Relying on sequential Monte Carlo (SMC) samples, we provide a closed-form approximation of solutions to the problem of choosing an optimal Galerkin projection or an optimal low–rank linear approximation. Numerical results obtained for a standard geophysical model show the gain brought by exploiting this posterior information for building a reduced model.

6.5. Combining analog method and ensemble data assimilation

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

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

Nowadays, ocean and atmosphere sciences face a deluge of data from spatial observations, in situ monitoring as well as numerical simulations. The availability of these different data sources offer new opportunities, still largely underexploited, to improve the understanding, modeling and reconstruction of geophysical dynamics. The classical way to reconstruct the space–time variations of a geophysical system from observations relies on data assimilation methods using multiple runs of the known dynamical model. This classical framework may have severe limitations including its computational cost, the lack of adequacy of the model with observed data, modeling uncertainties. In [60], we explore an alternative approach and develop a fully data—driven framework, which combines machine learning and statistical sampling to simulate the dynamics of complex system. As a proof concept, we address the assimilation of the chaotic Lorenz–63 model and imputation of missing data in multisite wind and rain time series. We demonstrate that a nonparametric sampler from a catalog of historical datasets, namely local linear regression, combined with a classical stochastic data assimilation scheme, the ensemble Kalman filter and the particular filter, reach state–of–the–art performances, without online evaluations of the physical model. The use of local regression instead of analog sampler allows to improve the performance of the filters.

6.6. Classification trees, functional data, applications in biology

Participants: Valérie Monbet, Audrey Poterie.

This is a collaboration with Jean–François Dupuy (INSA Rennes) and Laurent Rouvière (université de Haute Bretagne, Rennes).

Classification and discriminant analysis methods have grown in depths during the past 20 years. Fisher linear discriminant analysis (LDA) is the basic but standard approach. As the structure and dimension of the data becomes more complex in a wide range of applications, such as functional data, there is a need for more flexible nonparametric classification and discriminant analysis tools, especially when the ratio of learning sample size to number of covariates is low and the covariates are highly correlated and the covariance matrix is highly degenerated or when the large number of covariates are generally weak in predicting the class labels. For some data such as spectrometry data, only some parts of the observed curves are discriminant leading to groups of variables.

We proposed a classification tree based on groups of variables. Like usual tree-based methods, the algorithm partitions the feature space into M regions, by recursively performing binary splits. The main difference is that each split is based on groups of variables and the boundary between both classes is the hyperplane which minimizes the Bayes risk in the set generated by the selected group of variables. We demonstrate on several toy examples and real spectrometry data that the performances of the proposed tree groups algorithm are at least as good as the one of the standard CART algorithm and group Lasso logistic regression.

7. Bilateral Contracts and Grants with Industry

7.1. Bilateral grants with industry

See 4.1.

7.1.1. Hybrid indoor navigation — PhD project at CEA LETI

Participants: François Le Gland, Kersane Zoubert–Ousseni.

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

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

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

In some applications, real time estimation of the trajectory is not needed, and a post processing framework may provide a better estimation of this trajectory. In [23], we present and compare three different algorithms to improve a real time trajectory estimation. Actually, two different smoothing algorithms and the Viterbi algorithm are implemented and evaluated. These methods improve the regularity of the estimated trajectory by reducing switches between hypotheses.

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

Participants: François Le Gland, Audrey Cuillery.

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

After the introduction of MHT (multi–hypothesis tracking) techniques in the nineties, multitarget tracking has recently seen promising developpments with the introduction of new algorithms such as the PHD (probability hypothesis density) filter [55], [62] or the HISP (hypothesised filter for independent stochastic populations) filter [47]. These techniques provide a unified multitarget model in a Bayesian framework [59], which makes it possible to design recursive estimators of a *multitarget probability density*. Two main approaches can be used here: sequential Monte Carlo (SMC, also kown as particle filtering), and Gaussian mixture (GM). A third approach, based on discretizing the state–space in a possibly adaptive way, could also be considered despite its larger computational load. These methods are well studied and provide quite good results for *contact output* data, which correspond to regularly spaced measurements of targets with a large SNR (signal–to–noise ratio). Here, the data is processed (compared with a detection threshold) in each resolution cell of the sensor, so as to provide a list of detections at a given time instant. Among these methods, the HISP filter has the best performance/computational cost ratio.

However, these classical methods are unefficient for targets with a low SNR, e.g. targets in far range or small targets with a small detection probability. For such targets, preprocessing (thresholding) the data is not a good idea, and a much better idea is to feed a tracking algorithm with the raw *sensor output* data directly. These new methods [30] require a precise modeling of the sensor physics and a direct access to the radar (or the sonar) raw data, i.e. to the signal intensity level in each azimuth/range cell. Note that these new methods seem well suited to new types of sensors such as lidar, since manufacturers do not integrate a detection module and do provide raw images of the signal intensity level in each azimuth/range cell.

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

8. Partnerships and Cooperations

8.1. Regional initiatives

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

Participants: François Le Gland, Valérie Monbet.

January 2015 to December 2017.

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

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

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

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

8.2. National initiatives

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

Participant: Frédéric Cérou.

Inria contract ALLOC 9452 — January 2015 to December 2017.

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

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

8.2.2. Advanced Geophysical Reduced–Order Model Construction from Image Observations (GERONIMO) — ANR programme Jeunes Chercheuses et Jeunes Chercheurs

Participant: Patrick Héas.

Inria contract ALLOC 8102 — March 2014 to February 2018.

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

8.3. European initiatives

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

Participant: Mathias Rousset.

January 2014 to December 2019.

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

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

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

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

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

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

Participant: Valérie Monbet.

January 2016 to December 2018.

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

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

8.4. International initiatives

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

Participants: Ramatoulaye Dabo, Frédéric Cérou, François Le Gland.

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

The question here is to develop adaptive multilevel splitting algorithms for models that are commonly used in epidemiology, such as SIR (susceptible, infectious, recovered) models [32], or more complex compartmental models. A significant advantage of adaptive multilevel splitting is its robustness, since it does not require too much knowledge about the behavior of the system under study. An interesting challenge would be to understand how to couple the algorithm with numerically efficient simulation methods such as τ–leaping [42]. Complexity bounds and estimation error bounds could also be studied.

9. Dissemination

9.1. Promoting scientific activities

9.1.1. Scientific events organisation

As part of statistics semester of Labex Lebesgue, Valérie Monbet has co–organized the 3rd workshop on Stochastic Weather Generators, held in Vannes in May 2016. This workshop aimed at bringing together a wide range of researchers, practitioners, and graduate students whose work is related to the stochastic modelling of meteorological variables and stochastic weather generators. Stochastic weather generators give us ability to reliably predict climate-related risks by simulating sequences of daily weather and climate consistent with specific aspects of climate variability and change. The simulated sequences of meteorological variables (rainfall, wind, temperature,etc.) are typically used as inputs into complex environmental and ecosystem models. They have a wide range of applications in hydrology, agriculture and environmental management.

Within the programme *École d'été France Excellence* promoted by the French embassy in China, she has co–organized a two–weeks Summer School in Statistics, held in Rennes in June/July 2016. This initiative has offered Chinese students the opportunity to attend graduate courses in statistics, including practical and seminar sessions.

9.1.2. Participation in workshops, seminars, lectures, etc.

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

Frédéric Cérou has given an invited talk on the convergence of adaptive multilevel splitting at the RESIM 2016 workshop held in Eindhoven in March/April 2016. He has given, jointly with Mathias Rousset, a talk on a central limit theorem for adaptive multilevel splitting at the 2nd meeting on Adaptive Multilevel Splitting and Rare Events, an event of the MSMath ERC project held at CERMICS in Marne–la–Vallée, in June 2016.

Patrick Héas has given a talk on learning geophysical systems from images, at the seminar of ENS Rennes, in April 2016, and a talk on reduced modeling from partial observations, at the SIAM conference on Uncertainty Quantification, held in Lausanne, in April 2016.

François Le Gland has given a talk on marginalization for rare event simulation in switching diffusions at the RESIM 2016 workshop held in Eindhoven in March/April 2016, and at the probability and statistics seminar of LJK (laboratoire Jean Kuntzmann) in Grenoble, in June 2016.

Valérie Monbet has given a talk on time varying autoregressive models for multisite weather generators at the 3rd workshop on Stochastic Weather Generators, held in Vannes in May 2016. She has also given given a series of three lectures (including a lab session) at the *École d'été France Excellence*Summer School in Statistics, held in Rennes in June/July 2016.

9.1.3. Research administration

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

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

9.2. Teaching, supervision, thesis committees

9.2.1. Teaching

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

François Le Gland gives

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

Valérie Monbet gives several courses on data analysis, on time series, and on mathematical statistics, all at université de Rennes 1 within the master on statistics and econometrics. She is also the director of the master on statistics and econometry at université de Rennes 1.

9.2.2. Supervision

François Le Gland has been supervising one PhD student

- Alexandre Lepoutre, title: *Tracking and detection in Track–Before–Detect context using particle filtering*, université de Rennes 1, started in October 2010, defense held on October 5, 2016, funding: ONERA grant, co–direction: Olivier Rabaste (ONERA, Palaiseau).

Frédéric Cérou and François Le Gland are jointly supervising one PhD student

- Ramatoulaye Dabo, provisional title: *Rare event simulation in epidemiology*, université Assane Seck de Ziguinchor (Senegal) and université de Rennes 1, started in September 2015, expected defense in 2018, co–direction: Alassane Diedhiou (université Assane Seck de Ziguinchor).

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

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

François Le Gland is supervising two other PhD students

- Kersane Zoubert–Ousseni, provisional title: *Particle filters for hybrid indoor navigation with smartphones*, université de Rennes 1, started in December 2014, expected defense in 2017, funding: CEA grant, co–direction: Christophe Villien (CEA LETI, Grenoble),
- Audrey Cuillery, provisional title: *Bayesian tracking from raw data*, université du Sud Toulon Var, started in April 2016, expected defense in 2019, funding: CIFRE grant with DCNS, co–direction: Claude Jauffret (université du Sud Toulon Var) and Dann Laneuville (DCNS, Nantes).

Valérie Monbet is supervising two other PhD students

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

Mathias Rousset is supervising one PhD student

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

9.2.3. Thesis committees

François Le Gland has been a reviewer for the PhD theses of Tepmony Sim (Télécom ParisTech, advisors: Randal Douc and François Roueff) and Clément Walter (université Denis Diderot, Paris, advisor: Josselin Garnier).

Valérie Monbet has been a member of the committee for the HDR of Mathieu Emily (université de Haute Bretagne, Rennes).

Mathias Rousset has been a member of the committee for the PhD thesis of Zofia Trstanova (Inria Grenoble, EPI NANO-D, advisor: Stéphane Redon).

10. Bibliography

Major publications by the team in recent years

[1] G. Biau, F. Cérou, A. Guyader.*On the rate of convergence of the bagged nearest neighbor estimate*, in "Journal of Machine Learning Research", February 2010, vol. 11, p. 687–712, http://jmlr.csail.mit.edu/papers/v11/biau10a.html.

[2] G. Biau, F. Cérou, A. Guyader.*On the rate of convergence of the functional k–nearest neighbor estimates*, in "IEEE Transactions on Information Theory", April 2010, vol. IT–56, n^{o} 4, p. 2034–2040, http://dx.doi.org/10.1109/TIT.2010.2040857.

[3] G. Biau, F. Cérou, A. Guyader.*New insights into approximate Bayesian computation*, in "Annales de l'Institut Henri Poincaré, Probabilités et Statistiques", February 2015, vol. 51, n^{o} 1, p. 376–403, http://dx.doi.org/10.1214/13-AIHP590.

[4] F. Cérou, A. Guyader.*Nearest neighbor classification in infinite dimension*, in "ESAIM : Probability and Statistics", 2006, vol. 10, p. 340–355, http://dx.doi.org/10.1051/ps:2006014.

[5] F. Cérou, A. Guyader.*Adaptive multilevel splitting for rare event analysis*, in "Stochastic Analysis and Applications", March 2007, vol. 25, n^{o} 2, p. 417–443, http://dx.doi.org/10.1080/07362990601139628.

[6] F. Cérou, P. Del Moral, F. Le Gland, P. Lezaud.*Genetic genealogical models in rare event analysis*, in "ALEA, Latin American Journal of Probability and Mathematical Statistics", 2006, vol. 1, p. 181–203, Paper 01–08.

[7] A. Guyader, N. Hengartner, É. Matzner–Løber.*Simulation and estimation of extreme quantiles and extreme probabilities*, in "Applied Mathematics & Optimization", October 2011, vol. 64, n^{o} 2, p. 171–196, http://dx.doi.org/10.1007/s00245-011-9135-z.

[8] F. Le Gland, V. Monbet, V.–D. Tran.*Large sample asymptotics for the ensemble Kalman filter*, in "The Oxford Handbook of Nonlinear Filtering", Oxford, D. O. Crisan, B. L. Rozovskii (editors), Oxford University Press, 2011, chap. 22, p. 598–631.

[9] F. LE GLAND, N. OUDJANE.*A sequential algorithm that keeps the particle system alive*, in "Stochastic Hybrid Systems : Theory and Safety Critical Applications", Berlin, H. A. P. BLOM, J. LYGEROS (editors), Lecture Notes in Control and Information Sciences, Springer–Verlag, 2006, n° 337, p. 351–389, http://dx.doi.org/10.1007/11587392_11.

[10] C. MUSSO, N. OUDJANE, F. LE GLAND.*Improving regularized particle filters*, in "Sequential Monte Carlo Methods in Practice", New York, A. DOUCET, N. DE FREITAS, N. J. GORDON (editors), Statistics for Engineering and Information Science, Springer–Verlag, 2001, chap. 12, p. 247–271, http://dx.doi.org/10.1007/978-1-4757-3437-9_12.

Publications of the year

Doctoral Dissertations and Habilitation Theses

[11] A. LEPOUTRE.*Detection and tracking in Track-Before-Detect context using particle filtering*, Université de Rennes 1, October 2016, https://hal.inria.fr/tel-01423238.

Articles in International Peer-Reviewed Journal

[12] J.-D. ALBERT, V. MONBET, A. JOLIVET-GOUGEON, N. FATIH, M. LE CORVEC, M. SECK, F. CHARPENTIER, G. COIFFIER, C. BOUSSARD-PLÉDEL, B. BUREAU, P. GUGGENBUHL, O. LORÉAL.*A novel method for a fast diagnosis of septic arthritis using mid infrared and deported spectroscopy*, in "Joint Bone Spine", 2016, vol. 83, n° 3, p. 318-323 [*DOI :* 10.1016/J.JBSPIN.2015.05.009], https://hal-univ-rennes1.archives-ouvertes.fr/hal-01243032.

[13] J. BESSAC, P. AILLIOT, J. CATTIAUX, V. MONBET.*Comparison of hidden and observed regime-switching autoregressive models for (u,v)-components of wind fields in the Northeast Atlantic*, in "Advances in Statistical Climatology, Meteorology and Oceanography", 2016, vol. 2, n° 1, p. 1-16 [*DOI :* 10.5194/ASCMO-2-1-2016], https://hal.archives-ouvertes.fr/hal-01250353.

[14] P. BUI QUANG, C. MUSSO, F. LE GLAND.*Particle filtering and the Laplace method for target tracking*, in "IEEE Transactions on Aerospace and Electronic Systems", February 2016, vol. 52, n° 1, p. 350-366 [*DOI :* 10.1109/TAES.2015.140419], https://hal.archives-ouvertes.fr/hal-01314343.

[15] F. CÉROU, A. GUYADER.*Fluctuation analysis of adaptive multilevel splitting*, in "The Annals of Applied Probability : an official journal of the institute of mathematical statistics", December 2016, vol. 26, n° 6, p. 3319-3380 [*DOI :* 10.1214/16-AAP1177], https://hal.archives-ouvertes.fr/hal-01417241.

[16] P. HÉAS, A. DRÉMEAU, C. HERZET.*An Efficient Algorithm for Video Super-Resolution Based On a Sequential Model*, in "SIAM Journal on Imaging Sciences", 2016, vol. 9, n° 2, p. 537-572, 37 pages [*DOI :* 10.1137/15M1023956], https://hal.inria.fr/hal-01158551.

[17] D. JACQUEMART, J. MORIO.*Adaptive interacting particle system algorithm for aircraft conflict probability estimation*, in "Aerospace Science and Technology", June 2016, vol. 55, p. 431-438 [*DOI :* 10.1016/J.AST.2016.05.027], https://hal.archives-ouvertes.fr/hal-01404144.

[18] M. LE CORVEC, F. CHARPENTIER, A. KACHENOURA, S. BENSAID, S. HENNO, E. BARDOU-JACQUET, B. TURLIN, V. MONBET, L. SENHADJI, O. LORÉAL, O. SIRE, J. F. BETAGNE, H. TARIEL, F. LAINÉ.*Fast and Non-Invasive Medical Diagnostic Using Mid Infrared Sensor: The AMNIFIR Project*, in "IRBM", 2016,

vol. 37, n° 2, p. 116-123 [*DOI :* 10.1016/J.IRBM.2016.03.003], https://hal-univ-rennes1.archives-ouvertes.fr/hal-01296780.

[19] A. LEPOUTRE, O. RABASTE, F. LE GLAND.*Multitarget likelihood for Track-Before-Detect applications with amplitude fluctuations*, in "IEEE Transactions on Aerospace and Electronic Systems", June 2016, vol. 52, n° 3, p. 1089-1107 [*DOI :* 10.1109/TAES.2016.140909], https://hal.archives-ouvertes.fr/hal-01393453.

[20] V. MONBET, P. AILLIOT.*Sparse vector Markov switching autoregressive models. Application to multivariate time series of temperature*, in "Computational Statistics and Data Analysis", 2017, vol. 108, p. 40-51 [*DOI :* 10.1016/J.CSDA.2016.10.023], https://hal.archives-ouvertes.fr/hal-01250058.

Articles in National Peer-Reviewed Journal

[21] J.-D. ALBERT, V. MONBET, A. JOLIVET-GOUGEON, N. FATIH, M. LE CORVEC, M. SECK, F. CHARPENTIER, G. COIFFIER, C. BOUSSARD-PLÉDEL, B. BUREAU, P. GUGGENBUHL, O. LORÉAL.*Une nouvelle méthode pour le diagnostic rapide d'arthrite septique utilisant la spectroscopie infrarouge*, in "Revue du Rhumatisme", 2016, vol. 83, p. 295–300 [*DOI :* 10.1016/J.RHUM.2016.05.007], https://hal-univ-rennes1.archives-ouvertes.fr/hal-01367198.

International Conferences with Proceedings

[22] P. HÉAS, C. HERZET.*Reduced-Order Modeling of Hidden Dynamics*, in "IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASPP)", Shangai, China, March 2016, p. 1268–1272 [*DOI :* 10.1109/ICASSP.2016.7471880], https://hal.inria.fr/hal-01246074.

[23] K. ZOUBERT-OUSSENI, C. VILLIEN, F. LE GLAND.*Comparison of post-processing algorithms for indoor navigation trajectories*, in "Proceedings of the 7th International Conference on Indoor Positioning and Indoor Navigation (IPIN)", Alcala de Henares, Spain, IEEE, October 2016, ISBN: 978-1-5090-2425-4nternational Conference on Indoor Positioning and Indoor Navigation [*DOI :* 10.1109/IPIN.2016.7743590], https://hal.inria.fr/hal-01423198.

Other Publications

[24] F. CÉROU, B. DELYON, A. GUYADER, M. ROUSSET.*A Central Limit Theorem for Fleming-Viot Particle Systems*, November 2016, Preprint Arxiv, https://hal.archives-ouvertes.fr/hal-01391689.

[25] C. HERZET, A. DRÉMEAU, P. HÉAS.*Model Reduction from Partial Observations*, 2016, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01394059.

[26] P. HÉAS, C. HERZET.*Low-rank Approximation and Dynamic Mode Decomposition*, 2016, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01394064.

[27] P. HÉAS, C. HERZET.*Optimal Low-Rank Dynamic Mode Decomposition*, January 2017, IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASPP), New Orleans, USA, 2017, https://hal.archives-ouvertes.fr/hal-01429975.

References in notes

[28] A. DOUCET, N. DE FREITAS, N. J. GORDON (editors). *Sequential Monte Carlo methods in practice*, Statistics for Engineering and Information Science, Springer–Verlag, New York, 2001, http://dx.doi.org/10.1007/978-1-4757-3437-9.

[29] M. S. ARULAMPALAM, S. MAKSELL, N. J. GORDON, T. CLAPP.*A tutorial on particle filters for online nonlinear / non–Gaussian Bayesian tracking*, in "IEEE Transactions on Signal Processing", February 2002, vol. SP–50, n^{o} 2 (Special issue on Monte Carlo Methods for Statistical Signal Processing), p. 174–188, http://dx.doi.org/10.1109/78.978374.

[30] M. BOCQUEL.*Random finite sets in multi–target tracking: efficient sequential MCMC implementation*, Department of Applied Mathematics, University of Twente, Enschede, October 2013, http://purl.utwente.nl/publications/87496.

[31] O. CAPPÉ, S. J. GODSILL, É. MOULINES.*An overview of existing methods and recent advances in sequential Monte Carlo*, in "Proceedings of the IEEE", May 2007, vol. 95, n^{o} 5 (Special issue on Large–Scale Dynamic Systems), p. 899–924, http://dx.doi.org/10.1109/JPROC.2007.893250.

[32] D. J. DALEY, J. GANI.*Epidemic modelling: An introduction*, Cambridge Studies in Mathematical Biology, Cambridge University Press, Cambridge, 1999, vol. 15, http://dx.doi.org/10.1017/CBO9780511608834.

[33] P.–T. DE BOER, D. P. KROESE, S. MANNOR, R. Y. RUBINSTEIN.*A tutorial on the cross–entropy method*, in "Annals of Operations Research", January 2005, vol. 134 (Special issue on the Cross-Entropy Method for Combinatorial Optimization, Rare Event Simulation and Neural Computation), n^{o} 1, p. 19–67, http://dx.doi.org/10.1007/s10479-005-5724-z.

[34] P. DEL MORAL.*Mean field simulation for Monte Carlo integration*, Monographs on Statistics and Applied Probability, Chapman & Hall / CRC Press, London, 2013, vol. 126, http://www.crcpress.com/product/isbn/9781466504059.

[35] P. DEL MORAL.*Feynman–Kac formulae. Genealogical and interacting particle systems with applications*, Probability and its Applications, Springer–Verlag, New York, 2004, http://dx.doi.org/10.1007/978-1-4684-9393-1.

[36] P. DEL MORAL, L. MICLO.*Branching and interacting particle systems approximations of Feynman–Kac formulae with applications to nonlinear filtering*, in "Séminaire de Probabilités XXXIV", Berlin, J. AZÉMA, M. ÉMERY, M. LEDOUX, M. YOR (editors), Lecture Notes in Mathematics, Springer–Verlag, 2000, vol. 1729, p. 1–145, http://dx.doi.org/10.1007/BFb0103798.

[37] R. DOUC, É. MOULINES.*Limit theorems for weighted samples with applications to sequential Monte Carlo methods*, in "The Annals of Statistics", October 2008, vol. 36, n^{o} 5, p. 2344–2376, http://dx.doi.org/10.1214/07-AOS514.

[38] E. B. DYNKIN, R. J. VANDERBEI.*Stochastic waves*, in "Transactions of the American Mathematical Society", February 1983, vol. 275, n^{o} 2, p. 771–779, http://dx.doi.org/10.2307/1999052.

[39] D. FOX, J. HIGHTOWER, L. LIAO, D. SCHULZ, G. BORRIELLO.*Bayesian filtering for location estimation*, in "IEEE Pervasive Computing", July/September 2003, vol. 2, n^{o} 3, p. 24–33, http://dx.doi.org/10.1109/MPRV.2003.1228524.

[40] D. FOX, S. THRUN, W. BURGARD, F. DELLAERT.*Particle filters for mobile robot localization*, in "Sequential Monte Carlo Methods in Practice", New York, A. DOUCET, N. DE FREITAS, N. J. GORDON (editors), Statistics for Engineering and Information Science, Springer–Verlag, 2001, chap. 19, p. 401–428, http://dx.doi.org/10.1007/978-1-4757-3437-9_19.

[41] D. FRENKEL, B. SMIT.*Understanding molecular simulation. From algorithms to applications*, Computational Science Series, 2nd, Academic Press, San Diego, 2002, vol. 1, http://www.sciencedirect.com/science/book/9780122673511.

[42] D. T. GILLESPIE.*Approximate accelerated stochastic simulation of chemically reacting systems*, in "The Journal of Chemical Physics", July 2001, vol. 115, n^o^ 4, p. 1716–1733, http://dx.doi.org/10.1063/1.1378322.

[43] P. GLASSERMAN.*Monte Carlo methods in financial engineering*, Applications of Mathematics, Springer–Verlag, New York, 2004, vol. 53, http://dx.doi.org/10.1007/978-0-387-21617-1.

[44] P. GLASSERMAN, P. HEIDELBERGER, P. SHAHABUDDIN, T. ZAJIC.*Multilevel splitting for estimating rare event probabilities*, in "Operations Research", July–August 1999, vol. 47, n^o^ 4, p. 585–600, http://dx.doi.org/10.1287/opre.47.4.585.

[45] N. J. GORDON, D. J. SALMOND, A. F. M. SMITH.*Novel approach to nonlinear / non–Gaussian Bayesian state estimation*, in "IEE Proceedings, Part F", April 1993, vol. 140, n^o^ 2, p. 107–113, http://dx.doi.org/10.1049/ip-f-2.1993.0015.

[46] F. GUSTAFSSON, F. GUNNARSSON, N. BERGMAN, U. FORSSELL, J. JANSSON, R. KARLSSON, P.–J. NORDLUND.*Particle filters for positioning, navigation, and tracking*, in "IEEE Transactions on Signal Processing", February 2002, vol. SP–50, n^o^ 2 (Special issue on Monte Carlo Methods for Statistical Signal Processing), p. 425–437, http://dx.doi.org/10.1109/78.978396.

[47] J. HOUSSINEAU.*Representation and estimation of stochastic populations*, School of Engineering and Physical Sciences, Heriot–Watt University, Edinburgh, August 2015, https://jeremiehoussineau.wordpress.com/ph-d-thesis-all-versions/.

[48] M. ISARD, A. BLAKE.CONDENSATION — *Conditional density propagation for visual tracking*, in "International Journal of Computer Vision", August 1998, vol. 29, n^o^ 1, p. 5–28, http://dx.doi.org/10.1023/A:1008078328650.

[49] G. KITAGAWA.*Monte Carlo filter and smoother for non–Gaussian nonlinear state space models*, in "Journal of Computational and Graphical Statistics", 1996, vol. 5, n^o^ 1, p. 1–25, http://www.jstor.org/stable/1390750.

[50] H. R. KÜNSCH.*Recursive Monte Carlo filters : Algorithms and theoretical analysis*, in "The Annals of Statistics", October 2005, vol. 33, n^o^ 5, p. 1983–2021, http://www.jstor.org/stable/3448632.

[51] P. L'ÉCUYER, V. DEMERS, B. TUFFIN.*Rare events, splitting, and quasi–Monte Carlo*, in "ACM Transactions on Modeling and Computer Simulation", April 2007, vol. 17, n^o^ 2 (Special issue honoring Perwez Shahabuddin), Article 9, http://dx.doi.org/10.1145/1225275.1225280.

[52] F. LE GLAND, N. OUDJANE.*A robustification approach to stability and to uniform particle approximation of nonlinear filters : the example of pseudo-mixing signals*, in "Stochastic Processes and their Applications", August 2003, vol. 106, n^o^ 2, p. 279-316, http://dx.doi.org/10.1016/S0304-4149(03)00041-3.

[53] F. LE GLAND, N. OUDJANE.*Stability and uniform approximation of nonlinear filters using the Hilbert metric, and application to particle filters*, in "The Annals of Applied Probability", February 2004, vol. 14, n^{o} 1, p. 144–187, http://dx.doi.org/10.1214/aoap/1075828050.

[54] J. S. LIU.*Monte Carlo strategies in scientific computing*, Springer Series in Statistics, Springer–Verlag, New York, 2001, http://dx.doi.org/10.1007/978-0-387-76371-2.

[55] R. P. S. MAHLER.*Multitarget Bayes filtering via first–order multitarget moments*, in "IEEE Transactions on Aerospace and Electronic Systems", October 2003, vol. AES–39, n^{o} 4, p. 1152–1178, http://dx.doi.org/10.1109/TAES.2003.1261119.

[56] B. RISTIĆ, M. S. ARULAMPALAM, N. J. GORDON.*Beyond the Kalman Filter : Particle Filters for Tracking Applications*, Artech House, Norwood, MA, 2004, http://www.artechhouse.com/international/books/945.aspx.

[57] R. Y. RUBINSTEIN, D. P. KROESE.*The cross–entropy method. A unified approach to combinatorial optimization, Monte Carlo simulation and machine learning*, Information Science and Statistics, Springer–Verlag, New York, 2004, http://dx.doi.org/10.1007/978-1-4757-4321-0.

[58] C. J. STONE.*Consistent nonparametric regression (with discussion)*, in "The Annals of Statistics", July 1977, vol. 5, n^{o} 4, p. 595–645, http://www.jstor.org/stable/2958783.

[59] L. D. STONE, R. L. STREIT, T. L. CORWIN, K. L. BELL.*Bayesian multiple target tracking*, 2nd, Artech House, Norwood, MA, 2014, http://www.artechhouse.com/international/books/2090.aspx.

[60] P. TANDÉO, P. AILLIOT, J. RUIZ, A. HANNART, B. CHAPRON, A. CUZOL, V. MONBET, R. EASTON, R. FABLET.*Combining analog method and ensemble data assimilation: application to the Lorenz–63 chaotic system*, in "Machine Learning and Data Mining Approaches to Climate Science", Cham, V. LAKSHMANAN, E. GILLELAND, A. MCGOVERN, M. TINGLEY (editors), Springer–Verlag, 2015, chap. 1, p. 3–12, http://dx.doi.org/10.1007/978-3-319-17220-0_1.

[61] S. THRUN, W. BURGARD, D. FOX.*Probabilistic robotics*, Intelligent Robotics and Autonomous Agents, The MIT Press, Cambridge, MA, 2005, http://mitpress.mit.edu/books/probabilistic-robotics.

[62] B.–N. VO, W.–K. MA.*The Gaussian mixture probability hypothesis density filter*, in "IEEE Transactions on Signal Processing", November 2006, vol. SP–54, n^{o} 11, p. 4091–4104, http://dx.doi.org/10.1109/TSP.2006.881190.

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

Project-Team CAIRN

Energy Efficient Computing ArchItectures

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

IN PARTNERSHIP WITH:
CNRS

Université Rennes 1

École normale supérieure de Rennes

RESEARCH CENTER
Rennes - Bretagne-Atlantique

THEME
Architecture, Languages and Compilation

Table of contents

Project-Team CAIRN

Creation of the Project-Team: 2009 January 01

Keywords:

Computer Science and Digital Science:

1.1. - Architectures
1.1.1. - Multicore
1.1.2. - Hardware accelerators (GPGPU, FPGA, etc.)
1.1.8. - Security of architectures
1.1.9. - Fault tolerant systems
1.1.10. - Reconfigurable architectures
1.1.12. - Non-conventional architectures
1.2.5. - Internet of things
1.2.6. - Sensor networks
2.2. - Compilation
2.2.1. - Static analysis
2.2.4. - Parallel architectures
2.2.5. - GPGPU, FPGA, etc.
2.2.6. - Adaptive compilation
4.4. - Security of equipment and software
7.12. - Computer arithmetic

Other Research Topics and Application Domains:

4.5. - Energy consumption
4.5.1. - Green computing
4.5.2. - Embedded sensors consumption
6.2.2. - Radio technology
6.2.4. - Optic technology
6.6. - Embedded systems
8.1. - Smart building/home
8.1.1. - Energy for smart buildings
8.1.2. - Sensor networks for smart buildings

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

1. Members

Research Scientists

François Charot [Researcher, Inria, Rennes]
Olivier Sentieys [Team Leader, Senior Researcher, Inria, HDR]
Arnaud Tisserand [Senior Researcher, CNRS, Lannion, until Nov. 2016, HDR]
Tomofumi Yuki [Inria, Researcher, since Nov 2016]

Faculty Members

Emmanuel Casseau [Professor, Univ. Rennes I, ENSSAT, Lannion, HDR]
Daniel Chillet [Professor, Univ. Rennes I, ENSSAT, Lannion, HDR]

Steven Derrien [Professor, Univ. Rennes I, ISTIC, Rennes, HDR]
Cédric Killian [Associate Professor, Univ. Rennes I, IUT, Lannion]
Angeliki Kritikakou [Associate Professor, Univ. Rennes I, ISTIC, Rennes]
Patrice Quinton [Professor, ENS Rennes, Rennes, HDR]
Christophe Wolinski [Professor, Univ. Rennes I, Director of ESIR, Rennes, HDR]

Technical Staff

Philippe Quémerais [Research Engineer (20%), Univ. Rennes I, ENSSAT, Lannion, until Aug. 2016]
Arnaud Carer [Research Engineer (half time), Univ. Rennes I, Lannion]
Pierre Guilloux [Univ. Rennes I, Lannion]
Christophe Huriaux [Inria, Lannion]
Ali Hassan El Moussawi [Inria, Rennes]
Thomas Lefeuvre [Univ. Rennes I, Rennes]
Nicolas Simon [Univ. Rennes I, Rennes, until Oct. 2016]
Nicolas Estibals [Inria, Rennes, until Dec. 2016]
Raphael Bardoux [Inria, Lannion, until Jan. 2016]

PhD Students

Franck Bucheron [DGA, Rennes, from Oct. 2011 (half time)]
Gaël Deest [Univ. Rennes I, MENRT grant, Rennes, from Oct. 2013]
Rengarajan Ragavan [Univ. Rennes I, granted by FP7 FlexTiles, Lannion, from Oct. 2013]
Mai-Thanh Tran [Univ. Rennes I, granted by Brittany Region/CG22, Lannion, from Oct. 2013]
Xuan Chien Le [Inria, granted by Brittany Region/LTC, Lannion, from Oct. 2013]
Benjamin Barrois [Univ. Rennes I, MENRT grant, Lannion, from Oct. 2014]
Gabriel Gallin [CNRS, granted by CominLabs, from Oct. 2014]
Jiating Luo [Univ. Rennes I, granted by China Gov., from Nov. 2014]
Van Dung Pham [Inria, granted by CominLabs, Lannion, from Dec. 2014]
Baptiste Roux [Inria, granted by DGA and Inria, Rennes, from Oct. 2014]
Audrey Lucas [CNRS, granted by DGA-PEC, Lannion, from Jan. 2016]
Rafail Psiakis [Univ. Rennes I, MENRT grant, from Oct. 2015]
Simon Rokicki [Univ. Rennes I, granted by ENS Rennes, from Oct. 2015]
Aymen Gammoudi [Univ. Rennes I, Lannion, from Sep. 2015]
Mael Gueguen [Univ. Rennes I, MENRT grant, Rennes, from Nov. 2016]
Genevieve Ndour [CEA Leti, Grenoble, from May 2016]
Joel Ortiz Sosa [Inria, Lannion, from Oct. 2016]
Kleanthis Papachatzopoulos [Inria, Rennes, from Oct. 2016]
Tara Petric [Inria, Rennes, from Nov. 2016]
Nicolas Roux [Inria, Lannion, from Oct. 2016]
Florent Berthier [CEA Leti, Grenoble, until Oct. 2016]
Ali Hassan El-Moussawi [Univ. Rennes I, granted by FP7 Alma, Rennes, until Dec. 2016]
Jérémie Métairie [CNRS, granted by ANR Pavois, Lannion, until Apr. 2016]

Post-Doctoral Fellows

Ashraf El-Antably [Inria, Lannion, from May 2016]
Imen Fassi [Univ. Rennes I, Rennes]
Imran Wali [Univ. Rennes I, Lannion, from Sep. 2016]
Atef Dorai [Univ. Rennes I, ATER, ENSSAT, Lannion, from Sep. 2016]
Karim Bigou [Univ. Rennes I, ATER, IUT, Lannion, until Aug. 2016]
Benoit Lopez [Univ. Rennes I, ATER, ISTIC, Rennes, until Aug. 2016]

Administrative Assistants

Nadia Derouault [Inria, Rennes]
Angélique Le Pennec [Univ. Rennes I, ENSSAT, Lannion, until Aug. 2016]
Emilie Carquin [Univ. Rennes I, ENSSAT, Lannion, from June 2016]

2. Overall Objectives

2.1. Overall Objectives

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

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

The scientific goal of the CAIRN group is to research new hardware architectures for domain-specific SoCs, along with their associated design and compilation flows. We particularly focus on on-chip integration of specialized and reconfigurable accelerators. Reconfigurable architectures, whose hardware structure may be adjusted before or even during execution, originate from the possibilities opened up by Field Programmable Gate Arrays (FPGA) [64] and then by Coarse-Grain Reconfigurable Arrays (CGRA) [67], [81] [1]. Recent evolutions in technology and modern hardware systems confirm that reconfigurable systems are increasingly used in recent and future applications (see e.g. Intel/Altera or Xilinx/Zynq solutions). This architectural model has received a lot of attention in academia over the last two decades [71], and is now considered for industrial use in many application domains. One first reason is that the rapidly changing standards or applications require frequent device modifications. In many cases, software updates are not sufficient to keep devices on the market, while hardware redesigns remain too expensive. Second, the need to adapt the system to changing environments (e.g., wireless channel, harvested energy) is another incentive to use runtime dynamic reconfiguration. Moreover, with technologies at 28 nm and below, manufacturing problems strongly impact electrical parameters of transistors, and transient errors caused by particles or radiations also often appear during execution: error detection and correction mechanisms or autonomic self-control can benefit from reconfiguration capabilities.

As chip density increased, power or energy efficiency has become “the Grail” of all chip architects. With the end of Dennard scaling [76], multicore architectures are hitting the *utilisation wall* and the percentage of transistors in a chip that can switch at full frequency drops at a fast pace [69]. However, this unused portion of a chip also opens up new opportunities for computer architecture innovations. Building specialized processors or hardware accelerators can come with orders-of-magnitude gains in energy efficiency. Since from the beginning of CAIRN in 2009, we advocate the interest of heterogeneous multicores, in which general-purpose processors (GPPs) are integrated with specialized accelerators, especially when built on reconfigurable hardware, which provides the best trade-off between power, performance, cost and flexibility. During the period, it therefore turns out that the time has come for these heterogeneous manycore architectures.

> Standard multicore architectures enable flexible software on fixed hardware, whereas reconfigurable architectures make possible **flexible software on flexible hardware**.

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

3. Research Program

3.1. Panorama

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

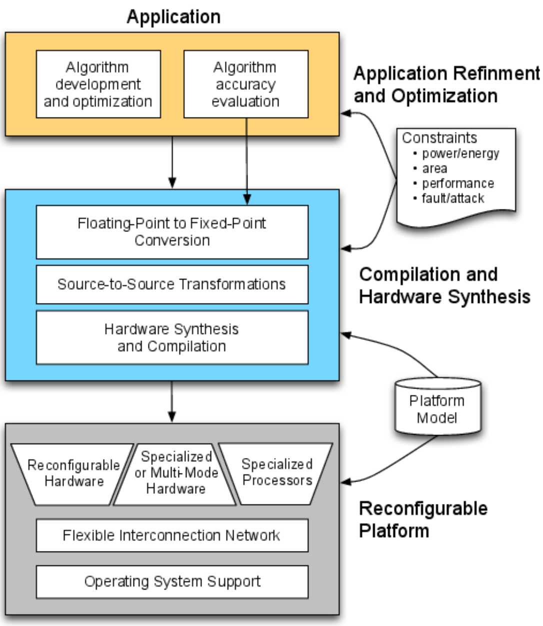


Figure 1. CAIRN's general design flow and related research themes

Figure 1 shows the global design flow we propose to develop. This flow is organized in levels which refer to our three research themes: application optimization (new algorithms, fixed-point arithmetic, advanced representations of numbers), architecture optimization (reconfigurable and specialized hardware, application-specific processors, arithmetic operators and functions), and stepwise refinement and code generation (code transformations, hardware synthesis, compilation).
In the rest of this part, we briefly describe the challenges concerning **new reconfigurable platforms** in Section 3.2 and the issues on **compiler and synthesis tools** related to these platforms in Section 3.3.

3.2. Reconfigurable Architecture Design

Nowadays, FPGAs are not only suited for application specific algorithms, but also considered as fully-featured computing platforms, thanks to their ability to accelerate massively parallelizable algorithms much faster than their processor counterparts [84]. They also support to be dynamically reconfigured. At runtime, partially reconfigurable regions of the logic fabric can be reconfigured to implement a different task, which allows for a better resource usage and adaptation to the environment. Dynamically reconfigurable hardware can also cope with hardware errors by relocating some of its functionalities to another, sane, part of the logic fabric. It could also provide support for a multi-tasked computation flow where hardware tasks are loaded on-demand at runtime. Nevertheless, current design flows of FPGA vendors are still limited by the use of one partial bitstream for each reconfigurable region and for each design. These regions are defined at design time and it is not possible to use only one bitstream for multiple reconfigurable regions nor multiple chips. The multiplicity of such bitstreams leads to a significant increase in memory. Recent research has been conducted in the domain of task relocation on a reconfigurable fabric. All of the related work was conducted on architectures from commercial vendors (e.g., Xilinx, Altera) which share the same limitations: the inner details of the bitstream are not publicly known, which limits applicability of the techniques. To circumvent this issue, most dynamic reconfiguration techniques are either generating multiple bitstreams for each location [66] or implementing an online filter to relocate the tasks [78]. Both of these techniques still suffer from memory footprint and from the online complexity of task relocation.

Increasing the level and grain of reconfiguration is a solution to counterbalance the FPGA penalties. Coarse-grained reconfigurable architectures (CGRA) provide operator-level configurable functional blocks and word-level datapaths [85], [72], [83]. Compared to FPGA, they benefit from a massive reduction in configuration memory and configuration delay, as well as for routing and placement complexity. This in turns results in an improvement in the computation volume over energy cost ratio, although with a loss of flexibility compared to bit-level operations. Such constraints have been taken into account in the design of DART[10], Adres [81] or polymorphous computing fabrics[12]. These works have led to commercial products such as the PACT/XPP [65] or Montium from Recore systems, without however a real commercial success yet. Emerging platforms like Xilinx/Zynq or Intel/Altera are about to change the game.

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

3.3. Compilation and Synthesis for Reconfigurable Platforms

In spite of their advantages, reconfigurable architectures, and more generally hardware accelerators, lack efficient and standardized compilation and design tools. As of today, this still makes the technology impractical for large-scale industrial use. Generating and optimizing the mapping from high-level specifications to reconfigurable hardware platforms are therefore key research issues, which have received considerable interest over the last years [70], [86], [82], [80], [79]. In the meantime, the complexity (and heterogeneity) of these platforms has also been increasing quite significantly, with complex heterogeneous multi-cores architectures becoming a *de facto* standard. As a consequence, the focus of designers is now geared toward optimizing overall system-level performance and efficiency [77]. Here again, existing tools are not well suited, as they fail at providing an unified programming view of the programmable and/or reconfigurable components implemented on the platform.

In this context, we have been pursuing our efforts to propose tools whose design principles are based on a tight coupling between the compiler and the target hardware architectures. We build on the expertise of the team members in High Level Synthesis (HLS) [6], ASIP optimizing compilers [13] and automatic parallelization for massively parallel specialized circuits [2]. We first study how to increase the efficiency of standard programmable processors by extending their instruction set to speed-up compute intensive kernels. Our focus is on efficient and exact algorithms for the identification, selection and scheduling of such instructions [7]. We address compilation challenges by borrowing techniques from high-level synthesis, optimizing compilers and automatic parallelization, especially when dealing with nested loop kernels. In addition, and independently of the scientific challenges mentioned above, proposing such flows also poses significant software engineering issues. As a consequence, we also study how leading edge software engineering techniques (Model Driven Engineering) can help the Computer Aided Design (CAD) and optimizing compiler communities prototyping new research ideas [14], [5], [3].

Efficient implementation of multimedia and signal processing applications (in software for DSP cores or as special-purpose hardware) often requires, for reasons related to cost, power consumption or silicon area constraints, the use of fixed-point arithmetic, whereas the algorithms are usually specified in floating-point arithmetic. Unfortunately, fixed-point conversion is very challenging and time-consuming, typically demanding up to 50% of the total design or implementation time. Thus, tools are required to automate this conversion. For hardware or software implementation, the aim is to optimize the fixed-point specification. The implementation cost is minimized under a numerical accuracy or an application performance constraint. For DSP-software implementation, methodologies have been proposed [8] to achieve fixed-point conversion. For hardware implementation, the best results are obtained when the word-length optimization process is coupled with the high-level synthesis [73]. Evaluating the effects of finite precision is one of the major and often the most time consuming step while performing fixed-point refinement. Indeed, in the word-length optimization process, the numerical accuracy is evaluated as soon as a new word-length is tested, thus, several times per iteration of the optimization process. Classical approaches are based on fixed-point simulations [74]. Leading to long evaluation times, they can hardly be used to explore the design space. Therefore, our aim is to propose closed-form expressions of errors due to fixed-point approximations that are used by a fast analytical framework for accuracy evaluation [11].

4. Application Domains

4.1. Panorama

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

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

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

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

5. Highlights of the Year

5.1. Highlights of the Year

Our work on accuracy evaluation and optimisation for fixed point arithmetic was presented during a tutorial "Fixed-point refinement, a guaranteed approach towards energy efficient computing" at HiPEAC Conference in January 2016 [60].
Members of CAIRN got six papers accepted at IEEE/ACM Design Automation and Test in Europe for 2017, one of the major events in design automation.

6. New Software and Platforms

6.1. Panorama

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

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

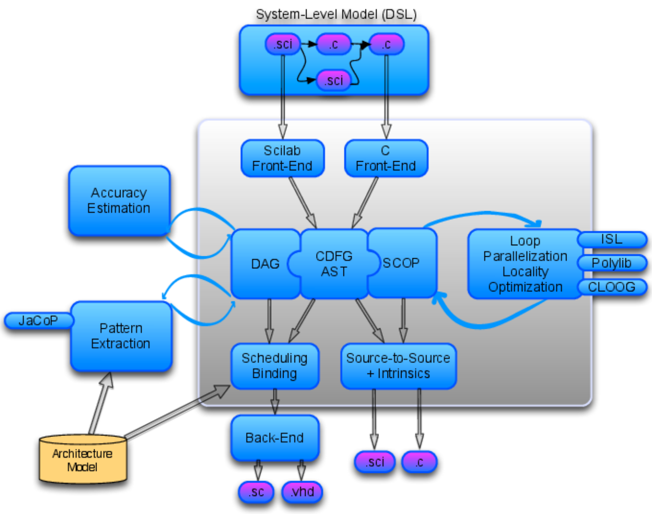


Figure 2. CAIRN's general software development framework.

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

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

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

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

6.2. Gecos

KEYWORDS: Source-to-source compiler - Model-driven software engineering - Retargetable compilation

SCIENTIFIC DESCRIPTION

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

Gecos is 100% Java based and takes advantage of modern model driven software engineering practices. It uses the Eclipse Modeling Framework (EMF) as an underlying infrastructure and takes benefits of its features to make it easily extensible. Gecos is open-source and is hosted on the Inria gforge at http://gecos.gforge.inria.fr.

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

FUNCTIONAL DESCRIPTION

Gecos provides a program transformation toolbox facilitating parallelisation of applications for heterogeneous multiprocessor embedded platforms. This includes a polyhedral loop transformation toolbox, efficient SIMD code generation for fixed point arithmetic data-types, coarse-grain parallelization engine targeting the dataflow actor model, and a Matlab/Scilab front-end. In addition to targeting programmable processors, Gecos can regenerate optimized code for High Level Synthesis tools.

- Participants: Steven Derrien, Nicolas Simon, Imen Fassi, and Ali Hassan El-Moussawi
- Partner: Université de Rennes 1
- Contact: Steven Derrien
- URL: http://gecos.gforge.inria.fr/doku/doku.php

6.3. ID-Fix

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

SCIENTIFIC DESCRIPTION

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

FUNCTIONAL DESCRIPTION

ID.Fix focuses on computational accuracy and can provide an optimised specification using fixed point arithmetic from a C source code with floating point data types. Fixed point arithmetic is very widely used in embedded systems as it provides better performance and is much more energy efficient. ID.Fix used an analytical model of the software code, which means it can explore more solutions and thereby produce much more efficient code than classical simulation-based tools.

- Participants: Olivier Sentieys, Benjamin Barrois and Nicolas Simon
- Partner: Université de Rennes 1
- Contact: Olivier Sentieys
- URL: http://idfix.gforge.inria.fr/doku.php

6.4. Zyggie

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

SCIENTIFIC DESCRIPTION

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

FUNCTIONAL DESCRIPTION

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

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

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

Figure 3. CAIRN*'s Ziggie platform for WBSN*

7. New Results

7.1. Reconfigurable Architecture Design

7.1.1. Dynamic Reconfiguration Support in FPGA

Participants: Olivier Sentieys, 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 developed 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.

In [4], we proposed a design flow for generating compressed configuration bitstreams abstracted from their final position on the logic fabric, the Virtual Bit-Streams (VBS). 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 VPR (Versatile Place and Route) framework was expanded to include bitstream generation features. The configuration stream format was proposed along with its associated decoding architecture. 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.

The emergence of 2.5D and 3D packaging technologies enables the integration of FPGA dice into more complex systems. Both heterogeneous manycore designs, which include an FPGA layer, and interposer-based multi-FPGA systems support the inclusion of reconfigurable hardware in 3D-stacked integrated circuits. In these architectures, the communication between FPGA dice or between FPGA and fixed-function layers often takes place through dedicated communication interfaces spread over the FPGA logic fabric, as opposed to an I/O ring around the fabric. In [39], we investigate the effect of organizing FPGA fabric I/O into coarse-grained interface blocks distributed throughout the FPGA fabric. Specifically, we consider the quality of results for the placement and routing phases of the FPGA physical design flow. We evaluate the routing of I/O signals of large applications through dedicated interface blocks at various granularities in the logic fabric, and study its implications on the critical path delay of routed designs. We show that the impact of such I/O routing is limited and can improve chip routability and circuit delay in many cases.

7.1.2. Hardware Accelerated Simulation of Heterogeneous Platforms

Participant: François Charot.

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

7.1.3. Optical Interconnections for 3D Multiprocessor Architectures

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

To address the issue of interconnection bottleneck in multiprocessor on a single chip, we study how an Optical Network-on-Chip (ONoC) can leverage 3D technology by stacking a specific photonics die. The objectives of this study target: i) the definition of a generic architecture including both electrical and optical components, ii) the interface between electrical and optical domains, iii) the definition of strategies (communication protocol) to manage this communication medium, and iv) new techniques to manage and reduce the power consumption of optical communications. The first point is required to ensure that electrical and optical components can be used together to define a global architecture. Indeed, optical components are generally larger than electrical components, so a trade-off must be found between the size of optical and electrical parts. For example, if the need in terms of communications is high, several waveguides and wavelengths must be necessary, and can lead to an optical area larger than the footprint of a single processor. In this case, a solution is to connect (through the optical NoC) clusters of processors rather than each single processor. For the second point, we study how the interface can be designed to take applications needs into account. From the different possible interface designs, we extract a high-level performance model of optical communications from losses induced by all optical components to efficiently manage Laser parameters. Then, the third point concerns the definition of high-level mechanisms which can handle the allocation of the communication medium for each data transfer between tasks. This part consists in defining the protocol of wavelength allocation. Indeed, the optical wavelengths are a shared resource between all the electrical computing clusters and are allocated at run time according to application needs and quality of service. The last point concerns the definition of techniques allowing to reduce the power consumption of on-chip optical communications. The power of each Laser can be dynamically tuned in the optical/electrical interface at run time for a given targeted bit-error-rate. Due to the relatively high power consumption of such integrated Laser, we study how to define adequate policies able to adapt the laser power to the signal losses.

We are currently designing an Optical-Network-Interface (ONI) to connect one processor, or a cluster of several processors, to the optical communication medium. This interface, constrained by the 10 Gb/s data-rate of the Lasers, integrates Error Correcting Codes and a communication manager. This manager can select, at run-time, the communication mode to use depending on timing or power constraints. Indeed, as the use of ECC is based on redundant bits, it increases the transmission time, but saves power for a given Bit Error Rate (BER). Moreover, our ONI allows for data to be sent using several wavelengths in parallel, hence increasing transmission bandwidth.

However, multiple signals sharing simultaneously a waveguide can lead to inter-channel crosstalk noise. This problem impacts the Signal to Noise Ratio (SNR) of the optical signal, which leads to an increase in the Bit Error Rate (BER) at the receiver side. In [40], [59], we proposed a Wavelength Allocation (WA) method allowing to search for performance and energy trade-offs based on application constraints. We showed that for a 16-core WDM ring-based ONoC architecture using 12 wavelengths, more than 100,000 allocation solutions exist and only 51 are on a Pareto front giving a tradeoff between execution time and energy per bit (derived from the BER). The optimized solutions reached reduce the execution time by 37% or the energy from 7,6fJ/bit to 4,4fJ/bit.

7.1.4. Communication-Based Power Modelling for Heterogeneous Multiprocessor Architectures

Participants: Baptiste Roux, Olivier Sentieys, Steven Derrien.

Programming heterogeneous multiprocessor architectures is a real challenge dealing with a huge design space. Computer-aided design and development tools try to circumvent this issue by simplifying instantiation mechanisms. However, energy consumption is not well supported in most of these tools due to the difficulty to obtain fast and accurate power estimation. To this aim, in [46] we proposed and validated a power model for such platforms. The methodology is based on micro-benchmarking to estimate the model parameters. The energy model mainly relies on the energy overheads induced by communications between processors in a parallel application. Power modelling and micro-benchmarks are validated using a Zynq-based heterogeneous architecture showing the accuracy of the model for several tested synthetic applications.

7.1.5. Arithmetic Operators for Cryptography and Fault-Tolerance

Participants: Arnaud Tisserand, Emmanuel Casseau, Pierre Guilloux, Karim Bigou, Gabriel Gallin, Audrey Lucas, Franck Bucheron, Jérémie Métairie.

Arithmetic Operators for Fast and Secure Cryptography.

Our paper [21], published in IEEE Transactions on Computers, extends our fast RNS modular inversion for finite fields arithmetic published at CHES 2013 conference. It is based on the binary version of the plus-minus Euclidean algorithm. In the context of elliptic curve cryptography (*i.e.* 160–550 bits finite fields), it significantly speeds-up modular inversions. In this extension, we propose an improved version based on both radix 2 and radix 3. This new algorithm leads to 30 % speed-up for a maximal area overhead about 4 % on Virtex 5 FPGAs. This work was done in the ANR PAVOIS project.

Our paper [32], presented at ARITH-23, presents an hybrid representation of large integers, or prime field elements, combining both positional and residue number systems (RNS). Our *hybrid position-residues* (HPR) number system mixes a high-radix positional representation and digits represented in RNS. RNS offers an important source of parallelism for addition, subtraction and multiplication operations. But, due to its non-positional property, it makes comparisons and modular reductions more costly than in a positional number system. HPR offers various trade-offs between internal parallelism and the efficiency of operations requiring position information. Our current application domain is asymmetric cryptography where HPR significantly reduces the cost of some modular operations compared to state-of-the-art RNS solutions. This work was done in the ANR PAVOIS project.

An ASIC circuit has been implemented in the 65nm ST CMOS technology and sent to fabrication in June 2016 (chip delivery is expected for January 2017). The implemented cryptoprocessor was designed for 256-bit prime finite fields elements and generic curves. It embeds: 1 multiplier, 1 adder and 1 inversion units for field-level computations. Various algorithms for scalar multiplication primitives can be programmed in software for curve-level computations. It was designed to evaluate algorithmic and arithmetic protections against side channel attacks (there is no hardware protection embedded in this ASIC version). This work was done in the ANR PAVOIS project.

In the HAH project, funded by CominLabs and Lebesgue Labex, we study hardware implementation of cryptoprocessors for hyperelliptic curves. The poster [61] presents the current state of the project for FPGA implementations.

Arithmetic Operators for Fault-Tolerance.

Various methods have been proposed for fault detection and fault tolerance in digital integrated circuits. In the case of *arithmetic circuits*, the selection of an efficient method depends on several elements: type of operation, type(s) of operand(s), computation algorithms, internal representations of numbers, optimizations at architecture and circuit levels, and acceptable accuracy level (i.e. mathematical error) of the result(s) including both rounding errors and errors due to the faults. High-level mathematical models are not sufficient to capture the effect of faults in arithmetic circuits. Simulation of intensive fault scenarios in all components of the arithmetic circuit (data-path, control, gates with important fan-out such as some partial products generation in large multipliers, etc.) is widely used. But cycle accurate and bit accurate software simulations at gate level are too slow for large circuits and numerous fault scenarios. *FPGA emulation* is a popular method to speed-up fault simulation.

We are developing an hardware-software platform dedicated to fault emulation for ASIC arithmetic circuits. The platform is based on a parallel cluster of Zynq FPGA cards and a Linux server. Various arithmetic circuits and fault models will be demonstrated in the context of digital signal and image processing. Our paper [57], presented at Compas, describes the very first version of our platform. This platform has also been presented in a poster at GDR SoC-SiP [58] and in a Demo Night at DASIP [56]. This work was done in the ANR ARDyT and Reliasic projects.

7.1.6. Adaptive Overclocking, Error Correction, and Voltage Over-Scaling for Error-Resilient Applications

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

Error detection and correction based on double-sampling is used as common technique to handle timing errors while scaling V_{dd} for energy efficiency. Implementation and advantages of double-sampling technique in FPGAs are simpler and significant compared to the conventional highly pipelined processors due to the higher flexibility of the reconfigurable architectures. It is common practice to insert shadow flipflop in the critical paths of the design, which will fail while scaling down the supply voltage, or to correct timing errors while over clocking the datapaths. Overclocking, and error detection and correction capabilities of these methods are limited due to the fixed speculation window used by these methods. In [44], we presented a Dynamic Speculation Window in double-sampling for timing error detection and correction in FPGAs. The proposed method employs online slack measurement and conventional shadow flipflop approach to adaptively overclock the design and also to detect and correct timing errors due to temperature and other variability effects. We demonstrated this method in the Xilinx VC707 Virtex 7 FPGA for various benchmarks. We achieved maximum of 71% overclocking for unsigned 32-bit multiplier with the area overhead of 1.9% LUTs and 1.7% FFs.

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

7.2. Compilation and Synthesis for Reconfigurable Platform

7.2.1. Adaptive dynamic compilation for low power embedded systems

Participants: Steven Derrien, Simon Rokicki.

Dynamic binary translation (DBT) consists in translating – at runtime – a program written for a given instruction set to another instruction set. Dynamic Translation was initially proposed as a means to enable code portability between different instruction sets and can be implemented in software or hardware. DBT is also used to improve the energy efficiency of high performance processors, as an alternative to out-of-order microarchitectures. In this context, DBT is used to uncover instruction level parallelism (ILP) in the binary program, and then target an energy efficient wide issue VLIW architecture. This approach is used in Transmeta Crusoe [75] and NVidia Denver [68] processors. Since DBT operates at runtime, its execution time is directly perceptible by the user, hence severely constrained. As a matter of fact, this overhead has often been reported to have a huge impact on actual performance, and is considered as being the main weakness of DBT based solutions. This is particularly true when targeting a VLIW processor: the quality of the generated code depends on efficient scheduling; unfortunately scheduling is known to be the most time-consuming component of a JIT compiler or DBT. Improving the responsiveness of such DBT systems is therefore a key research challenge. This is however made very difficult by the lack of open research tools or platform to experiment with such platforms. In this work, we have been addressing these two issues by developing an open hardware/software platform supporting DBT. The platform was designed using HLS tools and validated on a FPGA board. The DBT uses RISC-V as host ISA, and can target varying issue width VLIW architectures. Our platform uses custom hardware accelerators to improve the reactivity of our optimizing DBT flow. Our results show that, compared to a software implementation, our approach offers speed-up by $8\times$ while consuming $18\times$ less energy.

7.2.2. Leveraging Power Spectral Density for Scalable System-Level Accuracy Evaluation

Participants: Benjamin Barrois, Olivier Sentieys.

The choice of fixed-point word-lengths critically impacts the system performance by impacting the quality of computation, its energy, speed and area. Making a good choice of fixed-point word-length generally requires solving an NP-hard problem by exploring a vast search space. Therefore, the entire fixed-point refinement process becomes critically dependent on evaluating the effects of accuracy degradation. In [30], a novel technique for the system-level evaluation of fixed-point systems, which is more scalable and that renders better accuracy, was proposed. This technique makes use of the information hidden in the power-spectral density of quantization noises. It is shown to be very effective in systems consisting of more than one frequency sensitive components. Compared to state-of-the-art hierarchical methods that are agnostic to the quantization noise spectrum, we show that the proposed approach is $5\times$ to $500\times$ more accurate on some representative signal processing kernels.

7.2.3. Approximate Computing

Participants: Benjamin Barrois, Olivier Sentieys.

Many applications are error-resilient, allowing for the introduction of approximations in the calculations, as long as a certain accuracy target is met. Traditionally, fixed-point arithmetic is used to relax accuracy, by optimizing the bit-width. This arithmetic leads to important benefits in terms of delay, power and area. Lately, several hardware approximate operators were invented, seeking the same performance benefits. However, a fair comparison between the usage of this new class of operators and classical fixed-point arithmetic with careful truncation or rounding, has never been performed. In [31], we first compare approximate and fixed-point arithmetic operators in terms of power, area and delay, as well as in terms of induced error, using many state-of-the-art metrics and by emphasizing the issue of data sizing. To perform this analysis, we developed a design exploration framework, APXPERF, which guarantees that all operators are compared using the same operating conditions. Moreover, operators are compared in several classical real-life applications leveraging relevant metrics. In [31], we show that considering a large set of parameters, existing approximate adders and multipliers tend to be dominated by truncated or rounded fixed-point ones. For a given accuracy level and when considering the whole computation data-path, fixed-point operators are several orders of magnitude more accurate while spending less energy to execute the application. A conclusion of this study is that the entropy of careful sizing is always lower than approximate operators, since it require significantly less bits to be processed in the data-path and stored. Approximated data therefore always contain on average a greater amount of costly erroneous, useless information.

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

Participants: Daniel Chillet, Aymen Gammoudi.

Reconfigurable real-time embedded systems are constantly increasingly used in applications like autonomous robots or sensor networks. Since they are powered by batteries, these systems have to be energy-aware, to adapt to their environment and to satisfy real-time constraints. For energy harvesting systems, regular recharges of battery can be estimated, and by including this parameter in the operating system, it is then possible to develop strategy able to ensure the best execution of the application until the next recharge. In this context, operating system services must control the execution of tasks to meet the application constraints. Our objective concerns the proposition of a new real-time scheduling strategy that considers execution constraints such as the deadline of tasks and the energy.

To address this issue, we first focus on mono-processor scheduling [38] and propose to classify the tasks that have similar periods (or WCETs) in packs and to manage the execution parameters of these packs. For each reconfiguration scenario, parameter modifications are performed on packs/tasks to meet the real-time and energy constraints. Compared to previous work, task delaying is significantly improved in [36]. Furthermore, we also develop a strategy for multi-cores systems considering the dependencies between tasks [37] by adding the cost of communication between cores.

7.2.5. Optimization of loop kernels using software and memory information

Participant: Angeliki Kritikakou.

Current compilers cannot generate code that can compete with hand-tuned code in efficiency, even for a simple kernel like matrix–matrix multiplication (MMM). A key step in program optimization is the estimation of optimal values for parameters such as tile sizes and number of levels of tiling. The scheduling parameter values selection is a very difficult and time-consuming task, since parameter values depend on each other; this is why they are found by using searching methods and empirical techniques. To overcome this problem, the scheduling sub-problems must be optimized together, as one problem and not separately. In [24], an MMM methodology is presented where the optimum scheduling parameters are found by decreasing the search space theoretically, while the major scheduling sub-problems are addressed together as one problem and not separately according to the hardware architecture parameters and input size; for different hardware architecture parameters and/or input sizes, a different implementation is produced. This is achieved by fully exploiting the software characteristics (e.g., data reuse) and hardware architecture parameters (e.g., data caches sizes and associativities), giving high-quality solutions and a smaller search space. This methodology refers to a wide range of CPU and GPU architectures.

The size required to store an array is crucial for an embedded system, as it affects the memory size, the energy per memory access and the overall system cost. Existing techniques for finding the minimum number of resources required to store an array are less efficient for codes with large loops and not regularly occurring memory accesses. They have to approximate the accessed parts of the array leading to overestimation of the required resources. Otherwise their exploration time is increased with an increase over the number of the different accessed parts of the array. In [25], we propose a methodology to compute the minimum resources required for storing an array which keeps the exploration time low and provides a near-optimal result for regularly and non-regularly occurring memory accesses and overlapping writes and reads.

7.2.6. Adaptive Software Control to Increase Resource Utilization in Mixed-Critical Systems

Participant: Angeliki Kritikakou.

Automotive embedded systems need to cope with antagonist requirements: on the one hand, the users and market pressure push car manufacturers to integrate more and more services that go far beyond the control of the car itself. On the other hand, recent standardization efforts in the safety domain has led to the development of the ISO 26262 norm that defines means and requirements to ensure the safe operation of automotive embedded systems. In particular, it led to the definition of ASIL (Automotive Safety and Integrity Levels), i.e., it formally defines several criticality levels. Handling the increased complexity of new services makes new architectures, such as multi or many-cores, appealing choices for the car industry. Yet, these architectures provide a very low level of timing predictability due to shared resources, which goes in contradiction with timing guarantees required by ISO 26262. For highest criticality level tasks, Worst-Case Execution Time analysis (WCET) is required to guarantee that timing constraints are respected. The WCET analyzers consider the worst-case scenario: whenever a critical task accesses a shared resource in a multi/many-core platform, a WCET analyzer considers that all cores use the same resource concurrently. To improve the system performance, we proposed in a earlier work an approach where a critical task can be run in parallel with less critical tasks, as long as the real-time constraints are met. When no further interferences can be tolerated, the proposed run-time control in [54] suspends the low critical tasks until the termination of the critical task. In an automotive context, the approach can be translated as a highly critical partition, namely a classic AUTOSAR one, that runs on one dedicated core, with several cores running less critical Adaptive AUTOSAR application(s). We briefly describe in [54] the design of our proven-correct approach. Our strategy is based on a graph grammar to formally model the critical task as a set of control flow graphs on which a safe partial WCET analysis is applied and used at run-time to control the safe execution of the critical task.

8. Partnerships and Cooperations

8.1. Regional Initiatives

8.1.1. Images & Réseaux Competitivity Cluster - Embrace (2014-2016)

Participants: Raphaël Bardoux, Arnaud Carer, Olivier Sentieys.

Embrace (Embedded Radio Accelerator) is a project which involves CAIRN and two Small Medium Enterprises (SMEs): Digidia and PrimeGPS. Embrace aims at developing a software radio platform to enable the digital demodulation of HF signals. Both SMEs will use this platform as the first step to implement new products. These products will be dedicated to two different applications (Global Navigation Satellite System and Navigation Safety) at the heart of the markets of the SMEs. CAIRN goal is the technological transfer of the methods proposed by the team that enable the rapid prototyping of digital radios.

8.2. National Initiatives

8.2.1. ANR Blanc - PAVOIS (2012–2016)

Participants: Arnaud Tisserand, Emmanuel Casseau, Jérémie Métairie, Karim Bigou, Pierre Guilloux.

PAVOIS is a project on Arithmetic Protections Against Physical Attacks for Elliptic Curve based Cryptography that will provide novel implementations of curve based cryptographic algorithms on custom hardware platforms. A specific focus is placed on trade-offs between efficiency and robustness against physical attacks. It involves IRISA-CAIRN (Lannion) and LIRMM (Perpignan and Montpellier). Theoretical aspects include an investigation of how special number representations can be used to speed-up cryptographic algorithms, and protect cryptographic devices from physical attacks. On the practical side, we design innovative cryptographic hardware architectures of a specific processor based on the theoretical advancements described above to implement curve based protocols. For more details see http://pavois.irisa.fr.

8.2.2. ANR Ingénérie Numérique et Sécurité - ARDyT (2011-2016)

Participants: Arnaud Tisserand, Pierre Guilloux.

ARDyT is a project on a Reliable and Reconfigurable Dynamic Architecture. It involves IRISA-CAIRN (Lannion), Lab-STICC (Lorient), LIEN (Nancy) and ATMEL. The purpose of the ARDyT project is to provide a complete environment for the design of a fault tolerant and self-adaptable platform. Then, a platform architecture, its programming environment and management methodologies for diagnosis, testability and reliability have to be defined and implemented. The considered techniques are exempt from the use of hardened components for terrestrial and aeronautics applications for the design of low-cost solutions. For more details see http://ardyt.irisa.fr.

8.2.3. Labex CominLabs - BoWI (2012-2016)

Participants: Olivier Sentieys, Arnaud Carer.

The BoWi project (Body Wold Interactions) project aims at designing an accurate gesture and body movement estimation using very-small and low-power wearable sensor nodes, to propose pioneer interfaces for an emerging interacting world based on smart environments (house, media, information and entertainment systems...). Relying on Wireless Body Areas Sensor Networks, we propose an accurate Gesture and Body Movement estimation with extremely severe constraints in terms of footprint and energy consumption. The BoWI geolocation approach will combine radio communication distance measurement and inertial sensors and will also strongly benefit from cooperative techniques based on multiple observations and distributed computation. Different types of applications, such as health care, activity monitoring and environment control, are considered and prototyped. BoWI involves CAIRN, IRISA Granit (Lannion), IETR (Rennes), and Lab-STICC (Brest, Lorient, Vannes). For more details see http://www.bowi.cominlabs.ueb.eu.

8.2.4. Labex CominLabs - 3DCORE (2014-2018)

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

3DCORE (3D Many-Core Architectures based on Optical Network on Chip) is a project investigating new solutions based on silicon photonics to enhance by 2 to 3 magnitude orders energy efficiency and data rate of on-chip interconnect in the context of a many-core architecture. Moreover, 3DCore will take advantage of 3D technologies to design a specific optical layer suitable for a flexible and energy efficient high-speed optical network on chip (ONoC). 3DCORE involves CAIRN, FOTON (Rennes, Lannion) and Institut des Nanotechnologies de Lyon. For more details see http://www.3d-opt-many-cores.cominlabs.ueb.eu.

8.2.5. Labex CominLabs - RELIASIC (2014-2018)

Participants: Emmanuel Casseau, Arnaud Tisserand.

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

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

Participants: Arnaud Tisserand, Karim Bigou, Gabriel Gallin, Audrey Lucas.

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

8.3. European Initiatives

8.3.1. H2020 ARGO

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

Program: H2020-ICT-04-2015

Project acronym: ARGO

Project title: WCET-Aware Parallelization of Model-Based Applications for Heterogeneous Parallel Systems

Duration: Feb. 2016 - Feb. 2019

Coordinator: KIT

Other partners: KIT (DE), UR1/Inria/CAIRN (FR), Recore Systems (NL), TEI-WG (GR), Scilab Ent. (FR), Absint (DE), DLR (DE), Fraunhofer (DE)

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

8.3.2. ANR International ARTEFaCT

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

Program: ANR International France-Switzerland

Project acronym: ARTEFaCT

Project title: AppRoximaTivE Flexible Circuits and Computing for IoT

Duration: Feb. 2016 - Dec. 2019

Coordinator: CEA

Other partners: CEA-LETI (FR), CAIRN (FR), EPFL (SW)

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

8.4. International Initiatives

8.4.1. Inria Associate Teams

8.4.1.1. HARDIESSE

Title: Heterogeneous Accelerators for Reconfigurable DynamIc, Energy efficient, Secure SystEms

International Partner (Institution - Laboratory - Researcher):

University of Massachusetts at Ahmerst (United States) - Reconfigurable Computing Group - Russel Tessier

Start year: 2014

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

Rapid evolutions of applications and standards require frequent in-the-field system modifications and thus strengthens the need for adaptive devices. This need for a strong flexibility, combined with technology evolution (and the so-called power wall) has motivated the surge towards the use of multiple processor cores on a single chip (MPSoC). While it is now clear that we have entered the multi-core era, it is however indisputable that, especially for energy-efficient embedded systems, these architectures will have to be heterogeneous, by combining processor cores and specialized accelerators. We foresee a need for systems able to continuously adapt themselves to changing environments where software updates alone will not be enough for tackling energy management and error tolerance challenges. We believe that a dynamic and transparent adaptation of the hardware structure is the key to success. Security will also be an important challenge for embedded devices. Protections against physical attacks will have to be integrated in all secured components. In this Associated Team, we study new reconfigurable structures for such hardware accelerators with specific focus on: energy efficiency, runtime dynamic reconfiguration, security, and verification.

8.4.2. Inria International Partners

8.4.2.1. Declared Inria International Partners

8.4.2.1.1. LRS

Title: Loop unRolling Stones: compiling in the polyhedral model

International Partner (Institution - Laboratory - Researcher):

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

8.4.2.1.2. HARAMCOP

Title: Hardware accelerators modeling using constraint-based programming

International Partner (Institution - Laboratory - Researcher):

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

8.4.2.1.3. SPINACH

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

International Partner (Institution - Laboratory - Researcher):

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

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

8.4.2.2. Informal International Partners

Imec (Belgium), Fault-tolerant computing architectures.

Ecole Polytechnique Fédérale de Lausanne - EPFL (Switzerland), Optimization of embedded systems using fixed-point arithmetic, approximate computing.

Technical University of Madrid - UPM (Spain), Optimization of embedded systems using fixed-point arithmetic.

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

Department of Electrical and Computer Engineering, University of Patras (Greece), Wireless Sensor Networks, Worst-Case Execution Time, priority scheduling, loop transformations for memory optimizations.

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

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

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

8.5. International Research Visitors

8.5.1. Visits of International Scientists

Prof. Maciej Cieselski, University of Massachusetts, Amherst, US, for three weeks in July. This visit was partly funded by HARDIESSE Inria Associate Team.

Prof. Daniel Massicotte, Université du Québec à Trois-Rivières, CA, for three weeks in December. This visit was funded by ISTIC.

Maroua Gam, LabTim (Technologie Imagerie Médicale), Monastir, Tunisia, for one month in March.

8.5.2. Visits to International Teams

Angeliki Kritikakou visited University of Patras, Greece, for 1 week in November. This visit was funded by U. Rennes 1.

Patrice Quinton visited University of Massachusetts, Amherst, US, for 1 week in December. This visit was funded by HARDIESSE Inria Associate Team.

Tomofumi Yuki visited University of Arizona, US, in June.

8.5.2.1. Sabbatical programme

Casseau Emmanuel

Date: Aug 2016 - Jul 2017

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

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

9. Dissemination

9.1. Promoting Scientific Activities

9.1.1. Scientific Events Selection

9.1.1.1. General Chair, Scientific Chair

E. Casseau was General Co-Chair of DASIP, Conference on Design and Architectures for Signal and Image Processing, October 12-14, 2016.

S. Derrien was Co-Chair of WRC, 10th HiPEAC Workshop on Reconfigurable Computing, January 18-20, 2016 (co-located with HiPEAC 2016).

T. Yuki was Co-Chair of IMPACT, 6th International Workshop on Polyhedral Compilation Techniques, January 18-20, 2016 (co-located with HiPEAC 2016).

9.1.1.2. Chair of Conference Program Committees

O. Sentieys was Track Chair at IEEE NEWCAS.

9.1.1.3. Member of the Conference Program Committees

D. Chillet was member of the technical program committee of HiPEAC RAPIDO, HiPEAC WRC, MCSoC, DCIS, ComPAS, DASIP, LP-EMS, ARC.

S. Derrien was a member of technical program committee of IEEE FPL and ARC conferences and of WRC and Impact workshops.

O. Sentieys was a member of technical program committee of IEEE/ACM DATE, IEEE FPL, ACM ENSSys, ACM SBCCI, IEEE ReConFig, CROWNCOM, FSP, FPGA4GPC.

T. Yuki was a member of technical program committee of SC'16, The International Conference for High Performance Computing, Networking, Storage and Analysis.

9.1.2. Journal

9.1.2.1. Member of the Editorial Boards

D. Chillet is member of the Editor Board of Journal of Real-Time Image Processing (JRTIP).

O. Sentieys is member of the editorial board of Journal of Low Power Electronics and International Journal of Distributed Sensor Networks.

A. Tisserand is Associate Editor of IEEE Transactions on Computers. He is a member of the editorial board of the International Journal of High Performance Systems Architecture, Inderscience.

9.1.3. Invited Talks

O. Sentieys gave an invited talk at FETCH (École d'hiver Francophone sur les Technologies de Conception des Systèmes embarqués Hétérogènes), Villard-de-Lans, France, in January 2016 on "Approximate Computing and Flexible Circuits for the IoT".

T. Yuki gave a half-day lecture at EJCP 2016, École Jeunes Chercheurs en Programmation, Lille.

T. Yuki gave an invited talk at University of Arizona in June 2016 on "Optimizing Compilers in High-Level Synthesis".

9.1.4. Leadership within the Scientific Community

D. Chillet is member of the Board of Directors of Gretsi Association.

F. Charot, O. Sentieys and A. Tisserand are members of the steering committee of a CNRS spring school for graduate students on embedded systems architectures and associated design tools (ARCHI).

O. Sentieys and A. Tisserand are members of the steering committee of a CNRS spring school for graduate students on low-power design (ECOFAC).

A. Tisserand is co-organizer and president of scientific council of Seminar on Security of Embedded Electronic Systems (IRISA-DGA).

O. Sentieys is a member of the steering committee of the GDR SOC-SIP.

9.1.5. Scientific Expertise

O. Sentieys served as a jury member in the EDAA Outstanding Dissertations Award (ODA).

9.2. Teaching - Supervision - Juries

9.2.1. Teaching

E. Casseau: signal processing, 16h, ENSSAT (L3)
E. Casseau: low power design, 6h, ENSSAT (M1)
E. Casseau: real time design methodology, 24h, ENSSAT (M1)
E. Casseau: computer architecture, 36h, ENSSAT (M1)
E. Casseau: system on chip and verification, 10h, Master by Research (SISEA) and ENSSAT (M2)
E. Casseau: high level synthesis, 12h, Master by Research (SISEA) and ENSSAT (M2)
E. Casseau: advanced processor architectures, 25h, Univ. of Science and Tech. of Hanoi (M2)
S. Derrien: component and system synthesis, 20h, Master by Research (MRI ISTIC) (M2)
S. Derrien: computer architecture, 12h, ENS Rennes (L3)
S. Derrien: computer architecture, 24h, ISTIC(L3)
S. Derrien: introduction to operating systems, 8h, ISTIC(M1)
S. Derrien: embedded architectures, 48h, ISTIC(M1)
S. Derrien: high-level synthesis, 6h, ISTIC(M1)
S. Derrien: software engineering project, 40h, ISTIC(M1)
F. Charot: processor architecture, 25h Univ. of Science and Tech. of Hanoi (M1)
D. Chillet: embedded processor architecture, 20h, ENSSAT (M1)
D. Chillet: multimedia processor architectures, 24h, ENSSAT (M2)
D. Chillet: low-power digital CMOS circuits, 6h, Telecom Bretagne (M2)
C. Killian: digital electronics, 62h, IUT Lannion (L1)
C. Killian: signal processing, 36h, IUT Lannion (L2)
C. Killian: automated measurements, 56h, IUT Lannion (L2)
C. Killian: measurement chain, 35h, IUT Lannion (L2)
C. Killian: embedded systems programming, 12h, IUT Lannion (L2)
C. Killian: automatic control, 9h, IUT Lannion (L2)
A. Kritikakou: computer architecture 1, 50h, ISTIC, Univ. Rennes 1 (L3)
A. Kritikakou: computer architecture 2, 50h, ISTIC, Univ. Rennes 1 (L3)
A. Kritikakou: operating systems 1, 24h, ISTIC, Univ. Rennes 1 (L3)
A. Kritikakou: operating systems 2, 64h, ISTIC, Univ. Rennes 1 (L3)
A. Kritikakou: multitasking operating systems, 45h, ISTIC, Univ. Rennes 1 (M1)
O. Sentieys: digital signal processing, 40h, ENSSAT (M1)
O. Sentieys: VLSI integrated circuit design, 40h, ENSSAT(M1)
O. Sentieys: high level synthesis, 16h, Master by Research (SISEA) and ENSSAT (M2)
A. Tisserand: multiprocessor architectures, 20h, ENSSAT and Master by Research (SISEA) (M2)
C. Wolinski: computer architectures, 92h, ESIR (L3)
C. Wolinski: design of embedded systems, 48h, ESIR (M1)
C. Wolinski: signal, image, architecture, 26h, ESIR (M1)
C. Wolinski: programmable architectures, 10h, ESIR (M1)
C. Wolinski: component and system synthesis, 10h, Master by Research (MRI ISTIC) (M2)

9.2.2. Teaching Responsibilities

C. Wolinski is the Director of ESIR.

S. Derrien is the responsible of the first year of the Master of Computer Science at ISTIC since Sep. 2012.
O. Sentieys is responsible of the ”Embedded Systems” major of the SISEA Master by Research.
D. Chillet is the responsible of the ICT Master of University of Science and Technology of Hanoi.
C. Killian is the responsible of the second year of the Physical Measurement DUT at IUT of Lannion.

ENSSAT stands for *”École Nationale Supérieure des Sciences Appliquées et de Technologie”* and is an *”École d’Ingénieurs”* of the University of Rennes 1, located in Lannion.
ISTIC is the Electrical Engineering and Computer Science Department of the University of Rennes 1.
ESIR stands for *”École supérieure d’ingénieur de Rennes”* and is an *”École d’Ingénieurs”* of the University of Rennes 1, located in Rennes.

9.2.3. Supervision

PhD: Florent Berthier, Study and Design of an Ultra Low Power Asynchronous Core for Sensor Networks, Dec. 2016, O. Sentieys, E. Beigne.

PhD: Ali Hassan El-Moussawi, Performance/Accuracy Trade-Off in Automatic Parallelization for Embedded Many-Core Platforms, Dec. 2016, S. Derrien.

PhD: Jérémie Métairie, Reconfigurable Arithmetic Units for Secure Cryptoprocessors, May 2016, A. Tisserand, E. Casseau.

PhD in progress: Benjamin Barrois, Approximate Computing: a New Paradigm for Energy-Efficient Computing Architectures, Oct. 2014, O. Sentieys.

PhD in progress: Franck Bucheron, Secure Virtualization for Embedded Systems, Oct. 2011, A. Tisserand.

PhD in progress: Gaël Deest, Computing with Errors: Error-Tolerant Machine Code Generation for Unreliable Embedded Hardware, Oct. 2013, S. Derrien, O. Sentieys.

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

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

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

PhD in progress: Xuan Chien Le, Indirect Monitoring in Self-Powered Wireless Sensor Networks for Smart Grid and Building Automation, Oct. 2013, O. Sentieys, B. Vrigneau.

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

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

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

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

PhD in progress: Kleanthis Papachatzopoulos, Predictable and fault-tolerant multicore architecture, Oct. 2016, A. Kritikakou, O. Sentieys.

PhD in progress: Tara Petric, Approximate@runtime: Playing with accuracy at run-time for low-power flexible circuits in IoT nodes, Nov. 2016, T. Yuki, O. Sentieys.

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

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

PhD in progress: Rengarajan Ragavan, Ultra-Low Power Reconfigurable Architectures for Computing and Control in Wireless Sensor Networks, Oct. 2013, O. Sentieys, C. Killian.

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

PhD in progress: Baptiste Roux, Architectural Exploration of a Low-Power Flexible Radio Embedded on Drones, Oct. 2014, O. Sentieys, M. Gautier.

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

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

10. Bibliography

Major publications by the team in recent years

[1] R. DAVID, S. PILLEMENT, O. SENTIEYS. *Energy-Efficient Reconfigurable Processsors*, in "Low Power Electronics Design", C. PIGUET (editor), Computer Engineering, Vol 1, CRC Press, August 2004, chap. 20.

[2] S. DERRIEN, S. RAJOPADHYE, P. QUINTON, T. RISSET. *High-Level Synthesis of Loops Using the Polyhedral Model: The MMAlpha Software*, in "High-Level Synthesis From Algorithm to Digital Circuit", P. COUSSY, A. MORAWIEC (editors), Springer Netherlands, 2008, p. 215-230, http://dx.doi.org/10.1007/978-1-4020-8588-8.

[3] C. GUY, B. COMBEMALE, S. DERRIEN, J. STEEL, J.-M. JÉZÉQUEL. *On Model Subtyping*, in "8th European Conference on Modelling Foundations and Applications (ECMFA)", Kgs. Lyngby, Denmark, July 2012, http://hal.inria.fr/hal-00695034.

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

[5] J.-M. JÉZÉQUEL, B. COMBEMALE, S. DERRIEN, C. GUY, S. RAJOPADHYE. *Bridging the Chasm Between MDE and the World of Compilation*, in "Journal of Software and Systems Modeling (SoSyM)", October 2012, vol. 11, n° 4, p. 581-597 [*DOI :* 10.1007/s10270-012-0266-8], https://hal.inria.fr/hal-00717219.

[6] B. LE GAL, E. CASSEAU, S. HUET. *Dynamic Memory Access Management for High-Performance DSP Applications Using High-Level Synthesis*, in "IEEE Transactions on VLSI Systems", 2008, vol. 16, n° 11, p. 1454-1464.

[7] K. MARTIN, C. WOLINSKI, K. KUCHCINSKI, A. FLOCH, F. CHAROT. *Constraint Programming Approach to Reconfigurable Processor Extension Generation and Application Compilation*, in "ACM transactions on Reconfigurable Technology and Systems (TRETS)", June 2012, vol. 5, n° 2, p. 1-38, http://doi.acm.org/10.1145/2209285.2209289.

[8] D. MENARD, D. CHILLET, F. CHAROT, O. SENTIEYS. *Automatic Floating-point to Fixed-point Conversion for DSP Code Generation*, in "Proc. ACM/IEEE CASES", October 2002.

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

[10] S. PILLEMENT, O. SENTIEYS, R. DAVID.*DART: A Functional-Level Reconfigurable Architecture for High Energy Efficiency*, in "EURASIP Journal on Embedded Systems (JES)", 2008, p. 1-13.

[11] R. ROCHER, D. MÉNARD, O. SENTIEYS, P. SCALART.*Analytical Approach for Numerical Accuracy Estimation of Fixed-Point Systems Based on Smooth Operations*, in "IEEE Transactions on Circuits and Systems. Part I, Regular Papers", October 2012, vol. 59, n^o^ 10, p. 2326 - 2339 [*DOI :* 10.1109/TCSI.2012.2188938], http://hal.inria.fr/hal-00741741.

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

[13] C. WOLINSKI, K. KUCHCINSKI, E. RAFFIN.*Automatic Design of Application-Specific Reconfigurable Processor Extensions with UPaK Synthesis Kernel*, in "ACM Trans. on Design Automation of Elect. Syst.", 2009, vol. 15, n^o^ 1, p. 1–36, http://doi.acm.org/10.1145/1640457.1640458.

[14] S. WULIANG, B. COMBEMALE, S. DERRIEN, R. FRANCE.*Using Model Types to Support Contract-Aware Model Substitutability*, in "9th European Conference on Modelling Foundations and Applications (ECMFA)", Montpellier, France, P. VAN GORP, T. RITTER, L. ROSE (editors), LNCS, Springer-Verlag Berlin Heidelberg, 2013, vol. 7949, p. 118-133 [*DOI :* 10.1007/978-3-642-39013-5_9], http://hal.inria.fr/hal-00808770.

Publications of the year

Doctoral Dissertations and Habilitation Theses

[15] F. BERTHIER.*Design of an ultra low power processor for wireless sensor nodes*, Université de Rennes 1, France, December 2016, https://hal.inria.fr/tel-01423146.

[16] A. H. EL MOUSSAWI.*SIMD-aware Word Length Optimization for Floating-point to Fixed-point Conversion targeting Embedded Processors*, Universite de Rennes 1, December 2016, https://hal.inria.fr/tel-01425642.

[17] J. MÉTAIRIE.*Contributions to GF*(2^m) *arithmetic operators for elliptic curve cryptography*, Université Rennes 1, May 2016, https://hal.archives-ouvertes.fr/tel-01324924.

[18] J. MÉTAIRIE.*Contributions to GF(2m) Operators for Cryptographic Purposes*, Université Rennes 1, May 2016, https://tel.archives-ouvertes.fr/tel-01387919.

Articles in International Peer-Reviewed Journal

[19] M. M. ALAM, E. BEN HAMIDA, O. BERDER, O. SENTIEYS, D. MENARD.*A Heuristic Self-Adaptive Medium Access Control for Resource-Constrained WBAN Systems*, in "IEEE Access", April 2016, vol. 4, p. 1287-1300, https://hal.archives-ouvertes.fr/hal-01396104.

[20] F. BERTHIER, E. BEIGNE, F. HEITZMANN, O. DEBICKI, J.-F. CHRISTMANN, A. VALENTIAN, O. BILLOINT, E. AMAT, D. MORCHE, S. CHAIRAT, O. SENTIEYS.*UTBB FDSOI suitability for IoT applications: Investigations at device, design and architectural levels*, in "Solid-State Electronics", 2016, vol. 125, p. 14 - 24 [*DOI :* 10.1016/J.SSE.2016.09.003], https://hal.inria.fr/hal-01423144.

[21] K. BIGOU, A. TISSERAND.*Binary-Ternary Plus-Minus Modular Inversion in RNS*, in "IEEE Transactions on Computers", November 2016, vol. 65, n^o^ 11, p. 3495-3501 [*DOI :* 10.1109/TC.2016.2529625], https://hal.inria.fr/hal-01314268.

[22] R. BONAMY, S. BILAVARN, D. CHILLET, O. SENTIEYS.*Power Modeling and Exploration of Dynamic and Partially Reconfigurable Systems*, in "Journal of Low Power Electronics", September 2016, n^o^ September, https://hal.archives-ouvertes.fr/hal-01345664.

[23] M. FYRBIAK, S. ROKICKI, N. BISSANTZ, R. TESSIER, C. PAAR.*Hybrid Obfuscation to Protect against Disclosure Attacks on Embedded Microprocessors*, in "IEEE Transactions on Computers", 2017, https://hal.inria.fr/hal-01426565.

[24] V. KELEFOURAS, A. KRITIKAKOU, I. MPORAS, V. KOLONIAS.*A high performance Matrix-Matrix Multiplication Methodology for CPU and GPU architectures*, in "Journal of Supercomputing", 2016, p. 1-41 [*DOI :* 10.1007/s11227-015-1613-7], https://hal.archives-ouvertes.fr/hal-01255183.

[25] A. KRITIKAKOU, F. CATTHOOR, V. KELEFOURAS, C. GOUTIS.*Array Size Computation under Uniform Overlapping and Irregular Accesses*, in "ACM Transactions on Design Automation of Electronic Systems (TODAES)", 2016, https://hal.archives-ouvertes.fr/hal-01239705.

[26] P. A. M. OLIVEIRA, R. J. CINTRA, F. M. BAYER, S. KULASEKERA, A. MADANAYAKE.*Low-complexity Image and Video Coding Based on an Approximate Discrete Tchebichef Transform*, in "IEEE Transactions on Circuits and Systems for Video Technology", January 2016 [*DOI :* 10.1109/TCSVT.2016.2515378], https://hal.inria.fr/hal-01319513.

[27] L.-Q.-V. TRAN, A. DIDIOUI, C. BERNIER, G. VAUMOURIN, F. BROEKAERT, A. FRITCH.*Co-Simulating Complex Energy Harvesting WSN Applications: An In-Tunnel Wind Powered Monitoring Example*, in "International Journal of Sensor Networks", 2016, https://hal.inria.fr/hal-01264265.

[28] S. WANG, C. XIAO, W. LIU, E. CASSEAU.*A comparison of heuristic algorithms for custom instruction selection*, in "Microprocessors and Microsystems: Embedded Hardware Design (MICPRO)", August 2016, vol. 45, n^o^ A, 11, https://hal.inria.fr/hal-01354991.

Invited Conferences

[29] O. SENTIEYS, J. SEPÚLVEDA, S. LE BEUX, J. LUO, C. KILLIAN, D. CHILLET, I. O 'CONNOR, H. LI.*Design Space Exploration of Optical Interfaces for Silicon Photonic Interconnects*, in "2th International Workshop on Optical/Photonic Interconnects for Computing Systems (OPTICS Workshop), co-located with IEEE/ACM Design Automation and Test in Europe (DATE'16)", Dresden, Germany, March 2016, https://hal.inria.fr/hal-01293506.

International Conferences with Proceedings

[30] B. BARROIS, K. PARASHAR, O. SENTIEYS.*Leveraging Power Spectral Density for Scalable System-Level Accuracy Evaluation*, in "IEEE/ACM Conference on Design Automation and Test in Europe (DATE)", Dresden, Germany, March 2016, 6, https://hal.inria.fr/hal-01253494.

[31] B. BARROIS, O. SENTIEYS, D. MENARD.*The Hidden Cost of Functional Approximation Against Careful Data Sizing – A Case Study*, in "EEE/ACM Design Automation and Test in Europe (DATE)", Lausanne, France, 2017, https://hal.inria.fr/hal-01423147.

[32] K. BIGOU, A. TISSERAND.*Hybrid Position-Residues Number System*, in "ARITH: 23rd Symposium on Computer Arithmetic", Santa Clara, CA, United States, J. HORMIGO, S. OBERMAN, N. REVOL (editors), IEEE, July 2016, https://hal.inria.fr/hal-01314232.

[33] G. DEEST, N. ESTIBALS, T. YUKI, S. DERRIEN, S. RAJOPADHYE.*Towards Scalable and Efficient FPGA Stencil Accelerators*, in "6th International Workshop on Polyhedral Compilation Techniques (IMPACT'16), held with HIPEAC'16", Prague, Czech Republic, Proceedings of the IMPACT series, http://impact.gforge.inria.fr/, January 2016, https://hal.inria.fr/hal-01254778.

[34] A. H. EL MOUSSAWI, S. DERRIEN.*Superword Level Parallelism aware Word Length Optimization*, in "DATE - Design, Automation & Test in Europe Conference & Exhibition", Lausanne, Switzerland, D. ATIENZA, G. D. NATALE (editors), IEEE, March 2017, https://hal.inria.fr/hal-01425550.

[35] N. ESTIBALS, G. DEEST, A. EL-MOUSSAWI, S. DERRIEN.*System level synthesis for virtual memory enabled hardware threads*, in "Design, Automation & Test in Europe Conference & Exhibition", Dresden, France, March 2016, https://hal.inria.fr/hal-01424772.

[36] A. GAMMOUDI, A. BENZINA, M. KHALGUI, D. CHILLET.*New Reconfigurable Middleware for Adaptive RTOS in Ubiquitous Devices*, in "10th International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies", Venise, Italy, October 2016, https://hal.inria.fr/hal-01401716.

[37] A. GAMMOUDI, A. BENZINA, M. KHALGUI, D. CHILLET.*Real-Time Scheduling of Reconfigurable Battery-Powered Multi-Core Platforms*, in "28th International Conference on Tools with Artificial Intelligence", San Jose, United States, November 2016, https://hal.inria.fr/hal-01401712.

[38] A. GAMMOUDI, A. BENZINA, M. KHALGUI, D. CHILLET, A. GOUBAA.*Reconf-Pack: A Simulator for Reconfigurable Battery-Powered Real-Time Systems*, in "30th European Simulation and Modelling Conference", Las Palmas, Spain, October 2016, https://hal.inria.fr/hal-01401706.

[39] C. HURIAUX, O. SENTIEYS, R. TESSIER.*Effects of I/O Routing through Column Interfaces in Embedded FPGA Fabrics*, in "FPL - 26th International Conference on Field Programmable Logic and Applications", Lausanne, Switzerland, IEEE, August 2016, https://hal.inria.fr/hal-01341156.

[40] J. LUO, A. ELANTABLY, D. D. PHAM, C. KILLIAN, D. CHILLET, S. LE BEUX, O. SENTIEYS, I. O 'CONNOR.*Performance and Energy Aware Wavelength Allocation on Ring-Based WDM 3D Optical NoC*, in "Design, Automation & Test in Europe Conference & Exhibition (DATE) 2017", Lausanne, Switzerland, March 2017, https://hal.inria.fr/hal-01416958.

[41] T.-H. NGUYEN, P. SCALART, M. GAY, L. BRAMERIE, C. PEUCHERET, T. NGUYEN-TI, M. GAUTIER, O. SENTIEYS, J.-C. SIMON, M. JOINDOT.*Blind Adaptive Transmitter IQ Imbalance Compensation in M-QAM Optical Coherent Systems*, in "2016 IEEE International Conference on Communication (ICC 2016)", Kuala Lumpur, Malaysia, Communications (ICC), 2016 IEEE International Conference on, May 2016 [*DOI :* 10.1109/ICC.2016.7510925], https://hal.archives-ouvertes.fr/hal-01337225.

[42] T. H. NGUYEN, P. SCALART, M. GAY, L. BRAMERIE, C. PEUCHERET, O. SENTIEYS, J.-C. SIMON, M. JOINDOT.*Bi-harmonic decomposition-based maximum loglikelihood estimator for carrier phase estimation of coherent optical M-QAM*, in "Optical Fiber Communication Conference (OFC 2016)", Anaheim, CA, United States, Optical Fiber Communication Conference 2016, OSA (ISBN: 978-1-943580-07-1), March

2016, vol. DSP for Coherent Systems (Tu3K), Tu3K.3 [*DOI :* 10.1364/OFC.2016.TU3K.3], https://hal.archives-ouvertes.fr/hal-01309175.

[43] R. RAGAVAN, B. BARROIS, C. KILLIAN, O. SENTIEYS.*Pushing the Limits of Voltage Over-Scaling for Error-Resilient Applications*, in "Design, Automation & Test in Europe Conference & Exhibition (DATE 2017)", Lausanne, Switzerland, March 2017, https://hal.archives-ouvertes.fr/hal-01417665.

[44] R. RAGAVAN, C. KILLIAN, O. SENTIEYS.*Adaptive Overclocking and Error Correction Based on Dynamic Speculation Window*, in "ISVLSI", Pittsburgh, United States, 2016 IEEE Computer Society Annual Symposium on VLSI (ISVLSI), July 2016, p. 325 - 330 [*DOI :* 10.1109/ISVLSI.2016.13], https://hal.inria.fr/hal-01416945.

[45] S. ROKICKI, E. ROHOU, S. DERRIEN.*Hardware-Accelerated Dynamic Binary Translation*, in "IEEE/ACM Design, Automation & Test in Europe Conference & Exhibition (DATE)", Lausanne, Switzerland, March 2017, https://hal.inria.fr/hal-01423639.

[46] B. ROUX, M. GAUTIER, O. SENTIEYS, S. DERRIEN.*Communication-Based Power Modelling for Heterogeneous Multiprocessor Architecture*, in "IEEE 10th International Symposium on Embedded Multicore /Many-core Systems-on-Chip (MCSoC 2016)", Lyon, France, September 2016, https://hal.inria.fr/hal-01412835.

[47] M.-T. TRAN, M. GAUTIER, E. CASSEAU.*On the FPGA-based implementation of a flexible waveform from a high-level description: Application to LTE FFT case study*, in "EAI International Conference on Cognitive Radio Oriented Wireless Networks (Crowncom16)", Grenoble, France, May 2016, https://hal.inria.fr/hal-01302652.

National Conferences with Proceeding

[48] J. LUO, D. CHILLET, C. KILLIAN, S. LE BEUX, I. O 'CONNOR, O. SENTIEYS.*Crosstalk noise aware wavelength allocation in WDM 3D ONoC*, in "Colloque National du GDR SoC-SiP", Nantes, France, June 2016, https://hal.inria.fr/hal-01406355.

[49] J. LUO, D. CHILLET, C. KILLIAN, S. LE BEUX, I. O 'CONNOR, O. SENTIEYS.*Wavelength spacing optimization to reduce crosstalk in WDM 3D ONoC*, in "Conférence d'informatique en Parallélisme, Architecture et Système", Lorient, France, July 2016, https://hal.inria.fr/hal-01406341.

[50] V. D. PHAM, D. CHILLET, C. KILLIAN, S. LE BEUX, I. O 'CONNOR, O. SENTIEYS.*Gestion de la consommation d'un ONoC intégré dans un MPSoC*, in "Colloque National du GDR SoC-SiP", Nantes, France, June 2016, https://hal.inria.fr/hal-01414341.

[51] V. D. PHAM, C. KILLIAN, D. CHILLET, S. LE BEUX, O. SENTIEYS, I. O 'CONNOR.*Gestion de la consommation d'un réseau optique intégré dans un MPSoC*, in "Conférence d'informatique en Parallélisme, Architecture et Système", Lorient, France, July 2016, https://hal.inria.fr/hal-01406347.

[52] S. ROKICKI, E. ROHOU, S. DERRIEN.*Hybrid-JIT : Compilateur JIT Matériel/Logiciel pour les Processeurs VLIW Embarqués*, in "Confe´rence d'informatique en Paralle´lisme, Architecture et Syste`me (Compas)", Lorient, France, July 2016, https://hal.archives-ouvertes.fr/hal-01345306.

Conferences without Proceedings

[53] G. DEEST, N. ESTIBALS, T. YUKI, S. DERRIEN, S. RAJOPADHYE.*Towards Scalable and Efficient FPGA Stencil Accelerators*, in "IMPACT'16", Prague, Czech Republic, January 2016, https://hal.inria.fr/hal-01425018.

[54] A. KRITIKAKOU, T. MARTY, C. PAGETTI, C. ROCHANGE, M. LAUER, M. ROY.*Multiplexing Adaptive with Classic AUTOSAR? Adaptive Software Control to Increase Resource Utilization in Mixed-Critical Systems*, in "Workshop CARS 2016 - Critical Automotive applications : Robustness & Safety", Göteborg, Sweden, CARS 2016 - Critical Automotive applications : Robustness & Safety, September 2016, https://hal.archives-ouvertes.fr/hal-01375576.

Patents and standards

[55] F. BERTHIER, E. BEIGNE, F. HEITZMANN, O. DEBICKI, O. SENTIEYS.*Cœur de processeur asynchrone et microcontrôleur de nœud de capteur communicant comportant un tel cœur de processeur*, 2016, n^{o} 2016, https://hal.inria.fr/hal-01423133.

Other Publications

[56] P. GUILLOUX, A. TISSERAND.*Accurate Modeling of Fault Impact in Arithmetic Circuits*, October 2016, DASIP: Conference on Design and Architectures for Signal and Image Processing (Demo Night), Poster, https://hal.inria.fr/hal-01404772.

[57] P. GUILLOUX, A. TISSERAND.*Plateforme matérielle–logicielle d'émulation de fautes pour des opérateurs arithmétiques*, July 2016, 8, Compas 2016 : Conférence d'informatique en Parallélisme, Architecture et Système, https://hal.inria.fr/hal-01313051.

[58] P. GUILLOUX, A. TISSERAND.*Plateforme matérielle–logicielle à bas coût pour l'émulation de fautes*, June 2016, Colloque du GDR SoC-SiP, Poster, https://hal.inria.fr/hal-01346576.

[59] J. LUO, V. D. PHAM, C. KILLIAN, D. CHILLET, S. LE BEUX, I. O 'CONNOR, O. SENTIEYS.*POSTER: Wavelength Allocation for Efficient Communications on Optical Network-on-Chip*, October 2016, p. 1656 - 1658, Conference on Design and Architectures for Signal and Image Processing, Poster [*DOI :* 10.1145/2810103.2810122], https://hal.inria.fr/hal-01406328.

[60] O. SENTIEYS, D. MENARD, K. PARASHAR, D. NOVO.*Fixed-point refinement, a guaranteed approach towards energy efficient computing*, January 2016, Tutorial, https://hal.inria.fr/hal-01423184.

[61] A. TISSERAND, G. GALLIN.*Hardware and Arithmetic for Hyperelliptic Curves Cryptography*, November 2016, CominLabs Days 2016, Poster, https://hal.inria.fr/hal-01404755.

[62] M.-T. TRAN, E. CASSEAU, M. GAUTIER.*Demo abstract : FPGA-based implementation of a flexible FFT dedicated to LTE standard*, October 2016, 2, Conference on Design and Architectures for Signal and Image Processing (DASIP), Demo Night, Poster, https://hal.inria.fr/hal-01354992.

[63] Y. UGUEN, F. DE DINECHIN, S. DERRIEN.*Arithmetic Optimizations for High-Level Synthesis*, September 2016, working paper or preprint, https://hal.inria.fr/hal-01373954.

References in notes

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

[65] V. BAUMGARTE, G. EHLERS, F. MAY, A. NÜCKEL, M. VORBACH, M. WEINHARDT.*PACT XPP — A Self-Reconfigurable Data Processing Architecture*, in "The Journal of Supercomputing", 2003, vol. 26, n^{o} 2, p. 167–184.

[66] C. BECKHOFF, D. KOCH, J. TORRESEN.*Portable module relocation and bitstream compression for Xilinx FPGAs*, in "24th Int. Conf. on Field Programmable Logic and Applications (FPL)", 2014, p. 1–8.

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

[68] D. BOGGS, G. BROWN, N. TUCK, K. VENKATRAMAN.*Denver: NVIDIA's First 64-bit ARM Processor*, in "Micro", 2015.

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

[70] J. M. P. CARDOSO, P. C. DINIZ, M. WEINHARDT.*Compiling for reconfigurable computing: A survey*, in "ACM Comput. Surv.", June 2010, vol. 42, 13:1, http://doi.acm.org/10.1145/1749603.1749604.

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

[72] J. CONG, H. HUANG, C. MA, B. XIAO, P. ZHOU.*A Fully Pipelined and Dynamically Composable Architecture of CGRA*, in "IEEE Int. Symp. on Field-Program. Custom Comput. Machines (FCCM)", 2014, p. 9–16, http://dx.doi.org/10.1109/FCCM.2014.12.

[73] G. CONSTANTINIDES, P. CHEUNG, W. LUK.*Wordlength optimization for linear digital signal processing*, in "IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems", October 2003, vol. 22, n^{o} 10, p. 1432- 1442.

[74] M. COORS, H. KEDING, O. LUTHJE, H. MEYR.*Fast Bit-True Simulation*, in "Proc. ACM/IEEE Design Automation Conference (DAC)", Las Vegas, june 2001, p. 708-713.

[75] J. C. DEHNERT, B. K. GRANT, J. P. BANNING, R. JOHNSON, T. KISTLER, A. KLAIBER, J. MATTSON.*The Transmeta Code Morphing Software: Using Speculation, Recovery, and Adaptive Retranslation to Address Real-Life Challenges*, in "International Symposium on Code Generation and Optimization: Feedback-Directed and Runtime Optimization", 2003.

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

[77] A. HORMATI, M. KUDLUR, S. MAHLKE, D. BACON, R. RABBAH.*Optimus: efficient realization of streaming applications on FPGAs*, in "Proc. ACM/IEEE CASES", 2008, p. 41–50.

[78] H. KALTE, M. PORRMANN.*REPLICA2Pro: Task Relocation by Bitstream Manipulation in Virtex-II/Pro FPGAs*, in "3rd Conference on Computing Frontiers (CF)", 2006, p. 403–412.

[79] J.-E. LEE, K. CHOI, N. D. DUTT.*Compilation Approach for Coarse-Grained Reconfigurable Architectures*, in "IEEE Design and Test of Computers", 2003, vol. 20, n^{o} 1, p. 26-33, http://doi.ieeecomputersociety.org/10.1109/MDT.2003.1173050.

[80] H. LEE, D. NGUYEN, J.-E. LEE.*Optimizing Stream Program Performance on CGRA-based Systems*, in "52nd IEEE/ACM Design Automation Conference", 2015, p. 110:1–110:6, http://doi.acm.org/10.1145/2744769.2744884.

[81] B. MEI, S. VERNALDE, D. VERKEST, H. DE MAN, R. LAUWEREINS.*ADRES: An architecture with tightly coupled VLIW processor and coarse-grained reconfigurable matrix*, in "Proc. FPL", Springer, 2003, p. 61–70.

[82] N. R. MINISKAR, S. KOHLI, H. PARK, D. YOO.*Retargetable Automatic Generation of Compound Instructions for CGRA Based Reconfigurable Processor Applications*, in "Proc. ACM/IEEE CASES", 2014, p. 4:1–4:9, http://doi.acm.org/10.1145/2656106.2656125.

[83] Y. PARK, H. PARK, S. MAHLKE.*CGRA express: accelerating execution using dynamic operation fusion*, in "Proc. Int. Conf. on Compilers, Architecture, and Synthesis for Embedded Systems", New York, NY, USA, CASES'09, ACM, 2009, p. 271–280, http://doi.acm.org/10.1145/1629395.1629433.

[84] A. PUTNAM ET AL..*A reconfigurable fabric for accelerating large-scale datacenter services*, in "ACM/IEEE 41st International Symposium on Computer Architecture (ISCA)", June 2014, p. 13-24, http://dx.doi.org/10.1109/ISCA.2014.6853195.

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

[86] G. VENKATARAMANI, W. NAJJAR, F. KURDAHI, N. BAGHERZADEH, W. BOHM, J. HAMMES.*Automatic compilation to a coarse-grained reconfigurable system-on-chip*, in "ACM Trans. on Emb. Comp. Syst.", 2003, vol. 2, n^{o} 4, p. 560–589, http://doi.acm.org/10.1145/950162.950167.

Project-Team CELTIQUE

Software certification with semantic analysis

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

IN PARTNERSHIP WITH:
Université Rennes 1

École normale supérieure de Rennes

RESEARCH CENTER
Rennes - Bretagne-Atlantique

THEME
Proofs and Verification

Table of contents

Project-Team CELTIQUE

Creation of the Project-Team: 2009 July 01

Keywords:

Computer Science and Digital Science:

2.1. - Programming Languages
2.1.1. - Semantics of programming languages
2.1.2. - Object-oriented programming
2.1.3. - Functional programming
2.1.9. - Dynamic languages
2.2. - Compilation
2.2.1. - Static analysis
2.2.2. - Memory models
2.4. - Verification, reliability, certification
2.4.1. - Analysis
2.4.2. - Model-checking
2.4.3. - Proofs
4. - Security and privacy
4.5. - Formal methods for security

Other Research Topics and Application Domains:

6.1. - Software industry
6.1.1. - Software engineering
6.6. - Embedded systems

1. Members

Research Scientists

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

Faculty Members

Sandrine Blazy [Univ. Rennes I, Professor, HDR]
David Cachera [ENS Rennes, Associate Professor, HDR]
Delphine Demange [Univ. Rennes I, Associate Professor]
Thomas Genet [Univ. Rennes I, Associate Professor, HDR]
Serguei Lenglet [Univ. Lorraine, Associate Professor]
David Pichardie [ENS Rennes, Professor, HDR]
Charlotte Truchet [Univ. Nantes, Associate Professor]

Technical Staff

Laurent Guillo [CNRS, Senior engineer]
Pierre Wilke [Inria, until July 2016]

PhD Students

Oana Andreescu [Prove & Run, until Sep 2016, granted by CIFRE]
Martin Bodin [Inria, until Nov 2016]

Pauline Bolignano [Prove & Run, granted by CIFRE]
David Buhler [CEA]
Gurvan Cabon [Inria]
Alexandre Dang [Inria, from Oct 2016]
Yon Fernandez de Retana [Univ. Rennes I]
Julien Lepiller [Inria, from Oct 2016]
Florent Saudel [Amossys, granted by CIFRE]
Alix Trieu [Univ. Rennes I]
Yannick Zakowski [ENS Rennes]

Post-Doctoral Fellow

Cyrille Wiedling [Inria]

Visiting Scientist

Ahmad Salim Al-Sibahi [IT University of Copenhagen, from Sep 2016]

Administrative Assistant

Lydie Mabil [Inria]

Others

Aurele Barriere [Normale Sup Rennes, Internship, from May 2016 until Jul 2016]
Thomas Wood [Inria, PhD Student, from Oct 2016]

2. Overall Objectives

2.1. Project overview

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

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

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

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

3. New Software and Platforms

3.1. JSCert

Certified JavaScript
FUNCTIONAL DESCRIPTION

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

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

3.2. Javalib

FUNCTIONAL DESCRIPTION

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

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

3.3. SAWJA

Static Analysis Workshop for Java
KEYWORDS: Security - Software - Code review - Smart card
SCIENTIFIC DESCRIPTION

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

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

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

Version: 1.5

Programming language: Ocaml
FUNCTIONAL DESCRIPTION

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

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

3.4. Timbuk

KEYWORDS: Proof - Ocaml - Program verification - Tree Automata
FUNCTIONAL DESCRIPTION

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

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

3.5. CompCertSSA

KEYWORDS: Verified compilation - Single Static Assignment form - Optimization - Coq - OCaml

FUNCTIONAL DESCRIPTION

CompCertSSA is built on top of the C CompCert verified compiler, by adding a SSA-based middle-end (conversion to SSA, SSA-based optimizations, destruction of SSA).

Notably, the middle-end features:

- new important optimizations (Sparse Conditional Constant Propagation, and a coalescing phase on Conventional SSA)
- a generic dominance-based proof framework that rationalizes the proof process
- improved performance regarding compilation time

It is verified in the Coq proof assistant.

- Participant: Delphine Demange, David Pichardie, Yon Fernandez de Retana, Leo Stefanesco
- Contact: Delphine Demange
- URL: http://compcertssa.gforge.inria.fr/

4. New Results

4.1. Monitoring attacker knowledge with information flow analysis

Participants: Thomas Jensen, Frédéric Besson.

Motivated by the problem of stateless web tracking (fingerprinting) we have investigated a novel approach to hybrid information flow monitoring by tracking the knowledge that an attacker can learn about secrets during a program execution. We have proposed a general framework for combining static and dynamic information flow analysis, based on a precise representation of attacker knowledge. This hybrid analysis computes a precise description of what an attacker learns about the initial configuration (and in particular the secret part of it) by observing a specific output. An interesting feature of this knowledge-based information flow analysis is that it can be used to improve other information flow control mechanisms, such as no-sensitive upgrade. The whole framework is accompanied by a formalisation of the theory in the Coq proof assistant [18].

4.2. Semantic analysis of functional specifications of system software

Participants: Thomas Jensen, Oana Andreescu, Pauline Bolignano.

We have developed a static analysis for correlating input and output values in functional specifications, written in a functional, strongly typed, high-level specification formalism developed by the SME Prove & Run. In the context of interactive formal verification of complex systems, much effort is spent on proving the preservation of the system invariants. However, most operations have a localized effect on the system. Identifying correlations (in particular equalities) between input and output can substantially ease the proof burden for the programmer. Our correlation analysis is a flow-sensitive interprocedural analysis that handles arrays, structures and variant data types, and which computes a conservative approximation of the equality between sub-structures of input and of output fragments [27]. In a separate strand of work, we have used abstraction-based techniques for structuring and simplifying the proof of simulation between a high-level and a low-level specification of memory management algorithms in a hypervisor [22]. Both strands of work was carried out and validated on system software (a micro-kernel and a hypervisor) developed using the formal approach defined by Prove & Run.

4.3. Certified Static Analyses

4.3.1. Certified Semantics and Analyses for JavaScript

Participants: Martin Bodin, Gurvan Cabon, Thomas Jensen, Alan Schmitt.

We have continued our work on the certification of the semantics of JavaScript and of analyses for JavaScript on three different fronts.

First, on the language side, we have developed at tool in collaboration with Arthur Charguéraud (Inria Saclay) and Thomas Wood (Imperial College) to interactively explore the specification of JavaScript. More precisely, we have written a compiler for a subset of OCaml to a subset of JavaScript that generates an interpreter that can be executed step by step, inspecting both the state of the interpreted program but also the state of the interpreter. We have used this compiler on the JavaScript interpreter extracted from our Coq semantics of JavaScript. The resulting tool is available here and a demo can be run here. The tool has been presented to the Ecma TC39 committee in charge of standardizing JavaScript. We are currently identifying the improvements required to make it useful for the standardization process.

Second, Bodin, Schmitt, and Jensen have designed an abstract domain based on separation logic to faithfully abstract JavaScript heaps. This domain is able to capture interlinked dynamic and extensible objects, a central feature of the JavaScript memory model. In addition, we have introduced the notion of *membranes* that let us correctly define abstractions in a way that is compatible both with separation logic and abstract interpretation. As an extension of last year's work [32], this approach is globally correct as soon as each rule is independently proven correct. This result illustrates the robustness of our approach to define certified abstract semantics based on pretty-big-step semantics. This work has not yet been published.

Third, Cabon and Schmitt are developing a framework to automatically derive an information-flow tracking semantics from a pretty-big-step semantics. We have manually shown the approach works for complex examples, and are currently proving it in Coq. This work is submitted for publication.

4.3.2. Certified Analyses for C and lower-level programs

Participants: Sandrine Blazy, David Pichardie, Alix Trieu.

We have continued our work on the static analyzer Verasco [37], based on abstract interpretation and operating over most of the ISO C 1999 language (excluding recursion and dynamic allocation). Verasco establishes the absence of run-time errors in the analyzed programs. It enjoys a modular architecture that supports the extensible combination of multiple abstract domains. We have extended the memory abstract domain (that takes as argument any standard numerical abstract domain), so that it finely tracks properties about memory contents, taking into account union types, pointer arithmetic and type casts [19]. This memory domain is implemented and verified inside the Coq proof assistant with respect to the CompCert compiler memory model.

Motivated by applications to security and high efficiency, we are reusing the Verasco static analyzer and the CompCert compiler in order to design a lightweight and automated methodology for validating on low-level intermediate representations the results of a source-level static analysis. Our methodology relies on two main ingredients: a relative-safety checker, an instance of a relational verifier which proves that a program is safer than another, and a transformation of programs into defensive form which verifies the analysis results at runtime.

4.4. Certified Compilation

Participants: Sandrine Blazy, Frédéric Besson, Pierre Wilke, Alexandre Dang.

The COMPCERT C compiler provides the formal guarantee that the observable behaviour of the compiled code improves on the observable behaviour of the source code. A first limitation of this guarantee is that if the source code goes wrong, i.e. does not have a well-defined behaviour, any compiled code is compliant. Another limitation is that COMPCERT 's notion of observable behaviour is restricted to IO events.

Over the past years, we have developed the semantics theory so that unlike COMPCERT but like GCC, the binary representation of pointers can be manipulated much like integers and where memory is a finite resource. We have now a formally verified C compiler, COMPCERTS, which is essentially the COMPCERT compiler, albeit with a stronger formal guarantee. The semantics preservation theorem applies to a wider class of existing C programs and, therefore, their compiled version benefits from the formal guarantee of COMPCERTS. COMPCERTS preserves not only the observable behaviour of programs but also ensures that the memory consumption is preserved by the compiler. As a result, we have the formal guarantee that the compiled code requires no more memory than the source code. This ensures that the absence of stack-overflows is preserved by compilation.

The whole proof of COMPCERTS represents a significant proof-effort and the details can be found in Pierre Wilke's PhD thesis [39].

COMPCERTS also implements the Portable Software Fault Isolation approach pioneered by Kroll *et al.* [38]. The advantage of COMPCERTS is that the masking operation of pointers has a defined semantics and can therefore be directly reasoned about.

4.5. Mechanical Verification of SSA-based Compilation Techniques

Participants: Delphine Demange, Yon Fernandez de Retana, David Pichardie.

We have continued our work on the mechanical verification of SSA-based compilation techniques [30], [31], [36].

A crucial phase for efficient machine code generation is the destruction of a middle-end SSA-like IR. To this end, we have studied a variant of SSA, namely the Conventional SSA form, which simplifies the destruction back to non-SSA code (i.e. at the exit point of the middle-end). This had long remained a difficult problem, even in a non-verified environment. We formally defined and proved the properties of the generation of Conventional SSA. Finally, we implemented and proved correct a coalescing destruction of the Conventional SSA form, à la Boissinot et al. [33], where variables can be coalesced according to a refined notion of interference. Our CSSA-based, coalescing destruction allows us to coalesce more than 99% of introduced copies, on average, and leads to encouraging results concerning spilling and reloading during post-SSA allocation. This work has been published in [24].

4.6. Semantics for shared-memory concurrency

Participants: Gurvan Cabon, David Cachera, David Pichardie.

Modern multicore processor architectures and compilers of shared-memory concurrent programming languages provide only weak memory consistency guarantees. A *memory model* specifies which write action can be seen by a read action between concurrent threads.

In a previous work on the Java memory model [35], we defined in an axiomatic style, a memory model where we embed the reorderings of memory accesses directly in the semantics, so that formalizing optimizations and their correctness proof is easier.

This year, following a similar approach, we have studied the RMO (Relaxed- Memory Order) model. More precisely, we defined a new multibuffer operational semantics with write and read buffers. We also introduced an intermediate semantics inspired from Boudol et al. [34], where actions are reordered within a single pipeline. Finally, another model formalizes the reordering semantics in an axiomatic way. We fully proved the equivalence between the first two models and present a methodology for the remaining part. This work has been published in an international workshop [23].

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

Participant: Thomas Genet.

We develop a specific theory and the related tools for analyzing programs whose semantics is defined using term rewriting systems. The analysis principle is based on regular approximations of infinite sets of terms reachable by rewriting. Regular tree languages are (possibly) infinite languages which can be finitely represented using tree automata. To over-approximate sets of reachable terms, the tools we develop use the Tree Automata Completion (TAC) algorithm to compute a tree automaton recognizing a superset of all reachable terms. This over-approximation is then used to prove properties on the program by showing that some "bad" terms, encoding dangerous or problematic configurations, are not in the superset and thus not reachable. This is a specific form of, so-called, Regular Tree Model Checking. In [16], we have shown two results. The first result is a precision result guaranteeing that, for most of term rewriting systems known to have a regular set of reachable terms, TAC always compute it in an exact way. The second result shows that tree automata completion can be applied to functional programs to over-approximate their image. In particular, we have shown that tree automata completion computes a safe over-approximation of the image of any first-order, purely functional, complete and terminating program. Now, our first next objective is to demonstrate the accuracy of those regular approximations to perform lightweight formal verification of functional programs. The second objective is to lift those results to higher-order purely functional programs.

5. Partnerships and Cooperations

5.1. Regional Initiatives

5.1.1. Labex COMIN Labs Seccloud project

Participants: Frédéric Besson, Thomas Jensen, Alan Schmitt, Thomas Genet, Martin Bodin, Gurvan Cabon.

The SecCloud project, started in 2012, will provide a comprehensive language-based approach to the definition, analysis and implementation of secure applications developed using Javascript and similar languages. Our high level objectives is to enhance the security of devices (PCs, smartphones, ect.) on which Javascript applications can be downloaded, hence on client-side security in the context of the Cloud. We will achieve this by focusing on three related issues: declarative security properties and policies for client-side applications, static and dynamic analysis of web scripting programming languages, and multi-level information flow monitoring.

This is a joint project with Supelec Rennes and Ecole des Mines de Nantes.

5.2. National Initiatives

5.2.1. The ANR VERASCO project

Participants: Sandrine Blazy, Delphine Demange, David Pichardie.

Static program analysis, Certified static analysis

The VERASCO project (2012–06/2016) is funded by the call ISN 2011, a program of the Agence Nationale de la Recherche. It investigates the formal verification of static analyzers and of compilers, two families of tools that play a crucial role in the development and validation of critical embedded software. It is a joint project with the Inria teams ABSTRACTION, GALLIUM, The VERIMAG laboratory and the Airbus company.

5.2.2. The ANR AnaStaSec project

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

Static program analysis, Security, Secure compilation

The AnaStaSec project (2015–2018) aims at ensuring security properties of embedded critical systems using static analysis and security enhancing compiler techniques. The case studies are airborne embedded software with ground communication capabilities. The Celtique project focuses on software fault isolation which is a compiler technology to ensure by construction a strong segregation of tasks.

This is a joint project with the Inria teams ANTIQUE and PROSECCO, CEA-LIST, TrustInSoft, AMOSSYS and Airbus Group.

5.2.3. The ANR Binsec project

Participants: Frédéric Besson, Sandrine Blazy, Pierre Wilke, Julien Lepiller.

Binary code, Static program analysis

The Binsec project (2013–2017) is funded by the call ISN 2012, a program of the Agence Nationale de la Recherche. The goal of the BINSEC project is to develop static analysis techniques and tools for performing automatic security analyses of binary code. We target two main applicative domains: vulnerability analysis and virus detection.

Binsec is a joint project with the Inria CARTE team, CEA LIS, VERIMAG and EADS IW.

5.2.4. The ANR MALTHY project

Participant: David Cachera.

The MALTHY project, funded by ANR in the program INS 2013, aims at advancing the state-of-the-art in real-time and hybrid model checking by applying advanced methods and tools from linear algebra and algebraic geometry. MALTHY is coordinated by VERIMAG, involving CEA-LIST, Inria Rennes (Tamis and Celtique), Inria Saclay (MAXPLUS) and VISEO/Object Direct.

5.2.5. The ANR AJACS project

Participants: Martin Bodin, Gurvan Cabon, Thomas Jensen, Alan Schmitt.

The goal of the AJACS project is to provide strong security and privacy guarantees on the client side for web application scripts. To this end, we propose to define a mechanized semantics of the full JavaScript language, the most widely used language for the Web. We then propose to develop and prove correct analyses for JavaScript programs, in particular information flow analyses that guarantee no secret information is leaked to malicious parties. The definition of sub-languages of JavaScript, with certified compilation techniques targeting them, will allow us to derive more precise analyses. Finally, we propose to design and certify security and privacy enforcement mechanisms for web applications, including the APIs used to program real-world applications.

The project partners include the following Inria teams: Celtique, Indes, Prosecco, and Toccata; it also involves researchers from Imperial College as external collaborators. The project runs from December 2014 to June 2018.

5.2.6. The ANR DISCOVER project

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

The DISCOVER project project aims at leveraging recent foundational work on formal verification and proof assistants to design, implement and verify compilation techniques used for high-level concurrent and managed programming languages. The ultimate goal of DISCOVER is to devise new formalisms and proof techniques able to scale to the mechanized correctness proof of a compiler involving a rich class of optimizations, leading to efficient and scalable applications, written in higher-level languages than those currently handled by cutting-edge verified compilers.

In the light of recent work in optimizations techniques used in production compilers of high-level languages, control-flow-graph based intermediate representations seems too rigid. Indeed, the analyses and optimizations in these compilers work on more abstract representations, where programs are represented with data and control dependencies. The most representative representation is the sea-of-nodes form, used in the Java Hotspot Server Compiler, and which is the rationale behind the highly relaxed definition of the Java memory model. DISCOVER proposes to tackle the problem of verified compilation for shared-memory concurrency with a resolute language-based approach, and to investigate the formalization of adequate program intermediate representations and associated correctness proof techniques.

The project runs from October 2014 to September 2018.

5.3. European Initiatives

5.3.1. Collaborations in European Programs, Except FP7 & H2020

Program:CA COST Action CA15123

Project acronym: EUTYPES

Project title: European research network on types for programming and verification

Duration: 03/2016 to 03/2020

Coordinator: Herman Geuvers (Radboud University Nijmegen, The Netherlands)

Other partners: Austria, Belgium, Czech Republic, Denmark, Estonia, Finland, France, Macedonia, Germany, Hungary, Israel, Italy, Lithuania, Netherlands, Norway, Poland, Portugal, Romania, Serbia, Slovenia, Spain, Sweden, United Kingdom

Abstract: Types are pervasive in programming and information technology. A type defines a formal interface between software components, allowing the automatic verification of their connections, and greatly enhancing the robustness and reliability of computations and communications. In rich dependent type theories, the full functional specification of a program can be expressed as a type. Type systems have rapidly evolved over the past years, becoming more sophisticated, capturing new aspects of the behaviour of programs and the dynamics of their execution.

This COST Action will give a strong impetus to research on type theory and its many applications in computer science, by promoting (1) the synergy between theoretical computer scientists, logicians and mathematicians to develop new foundations for type theory, for example as based on the recent development of "homotopy type theory”, (2) the joint development of type theoretic tools as proof assistants and integrated programming environments, (3) the study of dependent types for programming and its deployment in software development, (4) the study of dependent types for verification and its deployment in software analysis and verification. The action will also tie together these different areas and promote cross-fertilisation.

5.4. International Initiatives

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

5.4.1.1. JCERT

Title: Verified Compilation of Concurrent Managed Languages

International Partner (Institution - Laboratory - Researcher):

Purdue University (United States) - School of Electrical and Computer Engineering (ECE) - Jan Vitek

Start year: 2014

See also: http://www.irisa.fr/celtique/ea/jcert/

Safety-critical applications demand rigorous, unambiguous guarantees on program correctness. While a combination of testing and manual inspection is typically used for this purpose, bugs latent in other components of the software stack, especially the compiler and the runtime system, can invalidate these hard-won guarantees. To address such concerns, additional laborious techniques such as manual code reviews of generated assembly code are required by certification agencies. Significant restrictions are imposed on compiler optimizations that can be performed, and the scope of runtime and operating system services that can be utilized. To alleviate this burden, the JCert project is implementing a verified compiler and runtime for managed concurrent languages like Java or C#.

5.4.2. Inria International Partners

5.4.2.1. WEBCERT

Title: Verified Trustworthy web Applications

International Partner (Institution - Laboratory - Researcher):

Imperial College (United Kingdom) - Department of Computing - Philippa Gardner

Duration: 2015 - 2019

Start year: 2015

See also: JSCert web page

The goal of the WebCert partnership is to extend the development of the JSCert formal semantics of JavaScript in the following domains: further mechanized specification, human-readable formal specification, program logic, verification tools, and the formalization of Defensive JavaScript.

5.4.2.2. Informal International Partners

Alan Schmitt is part of a Polonium Hubert Curien Partnership (PHC) with the University of Wrocław. This partnership is led by Sergueï Lenglet, from Loria, Nancy (currently visiting member of the Celtique project).

5.5. International Research Visitors

5.5.1. Visits of International Scientists

5.5.1.1. Internships

Thomas Wood

Date: Oct 2016 - Dec 2016

Institution: Imperial College (United Kingdom)

Ahmad Salim Al-Sibahi

Date: Sep 2016 - Jan 2017

Institution: IT University of Copenhagen (Denmark)

6. Dissemination

6.1. Promoting Scientific Activities

6.1.1. Scientific Events Organisation

6.1.1.1. General Chair, Scientific Chair

- PLMW@SPLASH 2016 (Programming Languages Mentoring Workshop) was chaired by Sandrine Blazy and Ulrik Prag-Schultz
- CoqPL 2017 (International Workshop on Coq for PL) was chaired by Sandrine Blazy and Emilio Jesus Gallego Arias

6.1.1.2. Member of the Organizing Committees

- JFLA 2016 (Journées Francophones des Langages Applicatifs) was locally organized by Julien Signoles and Alan Schmitt

6.1.2. Scientific Events Selection

6.1.2.1. Chair of Conference Program Committees

- VSTTE 2016 (Verified Software: Theories, Tools, and Experiments) was chaired by Sandrine Blazy and Marsha Chechik

6.1.2.2. Member of the Conference Program Committees

- CoqPL 2017 (International Workshop on Coq for PL) : Sandrine Blazy
- CPP 2017 (ACM SIGPLAN Conference on Certified Programs and Proofs) : Delphine Demange
- POPL 2017 (Symposium on Principles of Programming Languages) : Delphine Demange (External Program Committee)
- ESOP 2017 (European Symposium on Programming) : David Pichardie
- CC 2017 (International Conference on Compiler Construction) : David Pichardie
- IFL 2016 (International symposium on Implementation and application of Functional Languages) : Sandrine Blazy
- APLAS 2016 (Asian Symposium on Programming Languages and Systems) : Sandrine Blazy
- VSTTE 2016 (Verified Software: Theories, Tools, and Experiments) : Sandrine Blazy, Frédéric Besson
- GPCE 2016 (Generative Programming: Concepts & Experiences) : Sandrine Blazy
- DS@STAF 2016 (Doctoral Symposium) : Sandrine Blazy
- CPP 2016 (Certified Proofs and Programs) : Sandrine Blazy
- HaTT 2016 (International Workshop - Hammers for Type Theories) : Frédéric Besson
- AFADL 2016 (Approches Formelles dans l'Assistance au Développement de Logiciels) : Sandrine Blazy
- iFM 2016 (International Conference on integrated Formal Methods) : Delphine Demange
- FTfJP 2016 (Workshop on Formal Techniques for Java-like Programs) : Delphine Demange
- IFIP SEC 2016 (31st International Conference on ICT Systems Security and Privacy) : Thomas Jensen

6.1.2.3. Reviewer

- POPL 2017 (Symposium on Principles of Programming Languages): Alan Schmitt
- ESOP 2017 (European Symposium on Programming): Alan Schmitt
- VMCAI 2017 (International Conference on Verification, Model Checking, and Abstract Interpretation) : Delphine Demange

6.1.3. Journal

6.1.3.1. Reviewer - Reviewing Activities

- Journal of Software Evolution and Process: Sandrine Blazy
- International Journal of Computer Mathematics: Alan Schmitt
- Science of Computer Programming: Alan Schmitt

6.1.4. Invited Talks

- Journées nationales 2016 GDR Informatique Mathématique : Delphine Demange

6.1.5. Leadership within the Scientific Community

- Thomas Jensen is director of the Department NUMERIC of informatics, mathematics and electrical engineering at University Bretagne Loire.
- Thomas Jensen is leader of the security track of the LABEX Comin Labs.

6.1.6. Scientific Expertise

- Sandrine Blazy: expertise of 1 ANR project.
- Thomas Jensen: expertise of full project proposals for the ANR.

6.1.7. Research Administration

- Sandrine Blazy is member of Section 6 of the national committee for scientific research CoNRS from Sept. 2016.
- Sandrine Blazy is coordinator of the LTP (Languages, Types, Proofs) group of the French GDR GPL.

6.2. Teaching - Supervision - Juries

6.2.1. Teaching

Licence : Sandrine Blazy, Functional programming, 30h, L3, Université Rennes 1, France

Licence: Delphine Demange, Software Engineering, 40h, L2, Université de Rennes 1, France

Licence: Delphine Demange, Functional Programming, 75h, L1, Université de Rennes 1, France

Licence: Thomas Genet, Software Engineering, 58h, L2, Université de Rennes 1 / Istic, France

Licence : Alan Schmitt, Programmation Fonctionnelle, 72h (2 semestres), L3, Insa Rennes, France

Licence : David Pichardie, Algorithms, 36h, L3, ENS Rennes, France

Licence : David Cachera, Logic, 36h, L3, ENS Rennes, France

Master : Sandrine Blazy, Méthodes Formelles pour le développement de logiciels sûrs, 53h, M1, Université Rennes 1, France

Master : Thomas Genet, Formal Design and Verification, 108h, M1, Université de Rennes 1 / Istic, France

Master : Thomas Genet, Cryptographic Protocols, 24h, M2, Université de Rennes 1 / Istic, France

Master : David Pichardie, Mechanized Semantics, 15h, M2, Université Rennes 1, France

Master : Sandrine Blazy, Mechanized Semantics, 15h, M2, Université Rennes 1, France

Master : Sandrine Blazy, Semantics, 24h, M1, Université Rennes 1, France

Master : David Cachera, Semantics, 24h, M1, Université Rennes 1, France

Master : Sandrine Blazy, Software vulnerabilities, 20h, M2, Université Rennes 1, France

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

Master : Thomas Jensen, Program analysis and Software Security, 36h, M2, Université Rennes 1, France.

6.2.2. Supervision

PhD in progress : Alexandre Dang, Compiler for security, Octobre 2016, Thomas Jensen and Frédéric Besson

PhD in progress : Julien Lepiller, Binary Validation of Software Fault Isolation, Octobre 2016, Thomas Jensen and Frédéric Besson

PhD in progress : Gurvan Cabon, Analyse non locale certifiée en JavaScript grâce à une sémantique annotée, 1st september 2015, Alan Schmitt

PhD in progress : Florent Saudel, Vulnerability discovery, November 2015, Sandrine Blazy, Frédéric Besson and Dimitri Kirchner (Amossys)

PhD in progress : Alix Trieu, Formally verified compilation and static analysis, January 2016, Sandrine Blazy and David Pichardie

PhD in progress: David Bühler, Communication between analyses by deductive verification and abstract interpretation, November 2013, Sandrine Blazy and Boris Yakobowski (CEA)

PhD in progress : Yon Fernandez De Retana, Verified Optimising Compiler for high-level languages, 1st september 2015, David Pichardie and Delphine Demange

PhD in progress : Yannick Zakowski, Programs Logics for Concurrency, 1st september 2014, David Pichardie and David Cachera

PhD in progress : Oana Andreescu, Static analysis of functional specifications, 1st September 2013, Thomas Jensen, Stéphane Lescuyer (Prove & Run)

PhD in progress: Pauline Bolignano, Modeling and abstraction of system software, 1st November 2013, Thomas Jensen, Vincent Silés (Prove & Run)

Pierre Wilke, Formally verified compilation of low-level C code, Sandrine Blazy and Frédéric Besson, defended Nov 2016

Martin Bodin, Certified Analyses of JavaScript, Thomas Jensen and Alan Schmitt, defended Nov 2016

6.2.3. Juries

Sandrine Blazy, jury member (reviewer) for the PhD defense of Stefania Dumbrava, December 2016, Paris-Sud University, France

Sandrine Blazy, jury member (reviewer) for the PhD defense of Léon Gondelman, December 2016, Paris-Sud University, France

Sandrine Blazy, jury member (president) for the PhD defense of Thomas Degueule, December 2016, Rennes 1 University, France

Sandrine Blazy, jury member (president) for the PhD defense of Arjun Suresh, May 2016, Rennes 1 University, France

Sandrine Blazy, jury member for the selection of Inria CR (researcher) candidates, March and April 2016, Inria, Saclay, France.

Sandrine Blazy, jury member for the selection of a professeur at University of Perpignan, May 2016, Perpignan, France.

Alan Schmitt, jury member for the selection of Inria CR (researcher) candidates, March and April 2016, Inria, Rennes, France.

Delphine Demange, jury member for the selection of a Maître de Conférences at University Paris Diderot (Paris 7) / IRIF, May 2016, Paris, France.

Alan Schmitt, jury member (reviewer) for the PhD defense of Régis Spadotti, May 2016, Université Toulouse III

Alan Schmitt, jury member (reviewer) for the HDR defense of Nicolas Tabareau, November 2016, Université de Nantes

Thomas Jensen, jury member (reviewer) for the PhD defense of Denis Martinez, February 2016, Université de Montpellier

Thomas Jensen, jury member for the PhD defense of Oliver Schwarz, October 2016, KTH, Stockholm, Sweden

Thomas Jensen, jury member (reviewer) for the PhD defense of Rabah Laouadi, December 2016, Université de Montpellier

David Pichardie, jury member for the PhD defense of Jacques-Henri Jourdan, May 2016, Université de Paris Diderot

6.3. Popularization

Talk “Bug, Virus, Intrusion, Pirates... So many threats and no defense? Yes... maths.”, Thomas Genet, given three times in high schools close to Rennes.

7. Bibliography

Major publications by the team in recent years

[1] G. BARTHE, D. DEMANGE, D. PICHARDIE.*Formal Verification of an SSA-based Middle-end for CompCert*, in "ACM Transactions on Programming Languages and Systems (TOPLAS)", 2014, 35, https://hal.inria.fr/hal-01097677.

[2] F. BESSON, N. BIELOVA, T. JENSEN.*Hybrid Information Flow Monitoring Against Web Tracking*, in "CSF - 2013 IEEE 26th Computer Security Foundations Symposium", New Orleans, United States, 2013 [*DOI :* 10.1109/CSF.2013.23], http://hal.inria.fr/hal-00924138.

[3] F. BESSON, T. JENSEN, D. PICHARDIE.*Proof-Carrying Code from Certified Abstract Interpretation to Fixpoint Compression*, in "Theoretical Computer Science", 2006, vol. 364, n^{o} 3, p. 273–291.

[4] M. BODIN, A. CHARGUÉRAUD, D. FILARETTI, P. GARDNER, S. MAFFEIS, D. NAUDZIUNIENE, A. SCHMITT, G. SMITH.*A Trusted Mechanised JavaScript Specification*, in "POPL 2014 - 41st ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages", San Diego, United States, November 2013, http://hal.inria.fr/hal-00910135.

[5] B. BOYER, T. GENET, T. JENSEN.*Certifying a Tree Automata Completion Checker*, in "4th International Joint Conference, IJCAR 2008", Lectures Notes in Computer Science, Springer-Verlag, 2008, vol. 5195, p. 347–362.

[6] D. CACHERA, T. JENSEN, A. JOBIN, F. KIRCHNER.*Inference of polynomial invariants for imperative programs: a farewell to Gro¨bner bases*, in "Science of Computer Programming", 2014, vol. 93, 21 [*DOI :* 10.1016/J.SCICO.2014.02.028], https://hal.inria.fr/hal-00932351.

[7] D. CACHERA, T. JENSEN, D. PICHARDIE, V. RUSU.*Extracting a Data Flow Analyser in Constructive Logic*, in "Theoretical Computer Science", 2005, vol. 342, n^{o} 1, p. 56–78.

[8] D. DEMANGE, V. LAPORTE, L. ZHAO, D. PICHARDIE, S. JAGANNATHAN, J. VITEK.*Plan B: A Buffered Memory Model for Java*, in "Proc. of the 40th ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages, POPL 2013", Rome, Italy, ACM, 2013, http://hal.inria.fr/hal-00924716.

[9] T. GENET, V. RUSU.*Equational Approximations for Tree Automata Completion*, in "Journal of Symbolic Computation", 2010, vol. 45(5):574-597, May 2010, n^{o} 5, p. 574-597.

[10] L. HUBERT, T. JENSEN, V. MONFORT, D. PICHARDIE.*Enforcing Secure Object Initialization in Java*, in "15th European Symposium on Research in Computer Security (ESORICS)", Lecture Notes in Computer Science, Springer, 2010, vol. 6345, p. 101-115.

[11] J.-H. JOURDAN, V. LAPORTE, S. BLAZY, X. LEROY, D. PICHARDIE.*A formally-verified C static analyzer*, in "POPL 2015: 42nd ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages", Mumbai, India, ACM, January 2015, p. 247-259 [*DOI :* 10.1145/2676726.2676966], https://hal.inria.fr/hal-01078386.

Publications of the year

Articles in International Peer-Reviewed Journal

[12] A. Azevedo De Amorim, N. Collins, A. DeHon, D. Demange, C. Hriţcu, D. Pichardie, B. C. Pierce, R. Pollack, A. Tolmach.*A Verified Information-Flow Architecture*, in "Journal of Computer Security (JCS); Special Issue on Verified Information Flow Security", December 2016, vol. 24, n^o 6, p. 689–734, https://hal.archives-ouvertes.fr/hal-01424797.

[13] N. Bielova, F. Besson, T. Jensen.*Using JavaScript Monitoring to Prevent Device Fingerprinting*, in "ERCIM News", July 2016, https://hal.inria.fr/hal-01353997.

[14] S. Blazy, D. Bühler, B. Yakobowski.*Improving static analyses of C programs with conditional predicates*, in "Science of Computer Programming", March 2016, vol. 118, Extended version of the FMICS 2014 paper [*DOI :* 10.1145/2854065.2854082], https://hal.inria.fr/hal-01242077.

[15] S. Blazy, V. Laporte, D. Pichardie.*Verified Abstract Interpretation Techniques for Disassembling Low-level Self-modifying Code*, in "Journal of Automated Reasoning", 2016, vol. 56, n° 3, 26, Version étendue de l'article de la conférence ITP 2014 [*DOI :* 10.1007/s10817-015-9359-8], https://hal.inria.fr/hal-01243700.

[16] T. Genet.*Termination criteria for tree automata completion*, in "Journal of Logic and Algebraic Methods in Programming", 2016, vol. 85, Issue 1, part 1, p. 3-33 [*DOI :* 10.1016/j.jlamp.2015.05.003], https://hal.inria.fr/hal-01194533.

[17] F. Honsell, L. Liquori, P. Maksimovic, I. Scagnetto.*LLFP : A Logical Framework for modeling External Evidence, Side Conditions, and Proof Irrelevance using Monads*, in "Logical Methods in Computer Science", February 2016, https://hal.inria.fr/hal-01146059.

International Conferences with Proceedings

[18] F. Besson, N. Bielova, T. Jensen.*Hybrid Monitoring of Attacker Knowledge*, in "29th IEEE Computer Security Foundations Symposium", Lisboa, Portugal, 2016, https://hal.inria.fr/hal-01310572.

[19] S. Blazy, V. Laporte, D. Pichardie.*An Abstract Memory Functor for Verified C Static Analyzers*, in "ACM SIGPLAN International Conference on Functional Programming (ICFP 2016)", Nara, Japan, ACM, September 2016, 14 [*DOI :* 10.1145/2951913.2951937], https://hal.inria.fr/hal-01339969.

[20] S. Blazy, A. Trieu.*Formal Verification of Control-flow Graph Flattening*, in "Certified Proofs and Programs (CPP 2016)", Saint-Petersburg, United States, ACM (editor), Certified Proofs and Programs (CPP 2016), January 2016, 12, forthcoming [*DOI :* 10.1145/2854065.2854082], https://hal.inria.fr/hal-01242063.

[21] M. Bodin, T. Jensen, A. Schmitt.*An Abstract Separation Logic for Interlinked Extensible Records*, in "Vingt-septièmes Journées Francophones des Langages Applicatifs (JFLA 2016)", Saint-Malo, France, J. Signoles (editor), January 2016, https://hal.archives-ouvertes.fr/hal-01333600.

[22] P. Bolignano, T. Jensen, V. Siles.*Modeling and Abstraction of Memory Management in a Hypervisor*, in "Fundamental Approaches to Software Engineering (FASE'16)", Eindhoven, Netherlands, Proc. of Fundamental Approaches to Software Engineering (FASE'16), Springer, April 2016, vol. 9633, p. 214 - 230 [*DOI :* 10.1007/978-3-662-49665-7_13], https://hal.inria.fr/hal-01394174.

[23] G. CABON, D. CACHERA, D. PICHARDIE.*An Extended Buffered Memory Model With Full Reorderings*, in "FtFjp - Ecoop workshop", Rome, Italy, July 2016, p. 1 - 6 [*DOI :* 10.1145/2955811.2955816], https://hal.inria.fr/hal-01379514.

[24] D. DEMANGE, Y. FERNANDEZ DE RETANA.*Mechanizing conventional SSA for a verified destruction with coalescing*, in "25th International Conference on Compiler Construction", Barcelona, Spain, March 2016, https://hal.archives-ouvertes.fr/hal-01378393.

[25] C. FOURNET, C. KELLER, V. LAPORTE.*A Certified Compiler for Verifiable Computing*, in "IEEE 29th Computer Security Foundations Symposium, CSF 2016", Lisbonne, Portugal, June 2016, https://hal.inria.fr/hal-01397680.

[26] D. KÄSTNER, X. LEROY, S. BLAZY, B. SCHOMMER, M. SCHMIDT, C. FERDINAND.*Closing the Gap – The Formally Verified Optimizing Compiler CompCert*, in "SSS'17: Safety-critical Systems Symposium 2017", Bristol, United Kingdom, Proceedings of the Twenty-fifth Safety-Critical Systems Symposium, February 2017, https://hal.inria.fr/hal-01399482.

[27] A. OANA, T. JENSEN, S. LESCUYER.*Correlating Structured Inputs and Outputs in Functional Specifications*, in "Software Engineering and Formal Methods", Vienna, Austria, 14th Int. Software Engineering and Formal Methods conference, Springer, July 2016, vol. 9763, 19 [*DOI :* 10.1007/978-3-319-41591-8_7], https://hal.inria.fr/hal-01394178.

Conferences without Proceedings

[28] X. LEROY, S. BLAZY, D. KÄSTNER, B. SCHOMMER, M. PISTER, C. FERDINAND.*CompCert - A Formally Verified Optimizing Compiler*, in "ERTS 2016: Embedded Real Time Software and Systems, 8th European Congress", Toulouse, France, SEE, January 2016, https://hal.inria.fr/hal-01238879.

Books or Proceedings Editing

[29] S. BLAZY, M. CHÉCHIA (editors). *Verified Software: Theories, Tools, and Experiments - 8th International Conference, VSTTE 2016, Toronto, Canada, July 17-18, 2016. Proceedings*, Lecture Notes in Computer Science, Springer, Toronto, Canada, 2016, vol. 9971, https://hal.inria.fr/hal-01387207.

References in notes

[30] G. BARTHE, D. DEMANGE, D. PICHARDIE.*Formal Verification of an SSA-based Middle-end for CompCert*, in "ACM Transactions on Programming Languages and Systems (TOPLAS)", 2014, 35, https://hal.inria.fr/hal-01097677.

[31] S. BLAZY, D. DEMANGE, D. PICHARDIE.*Validating Dominator Trees for a Fast, Verified Dominance Test*, in "Proc. of the 6th International Conference on Interactive Theorem Proving (ITP 2015)", LNCS, Springer, 2015.

[32] M. BODIN, T. JENSEN, A. SCHMITT.*Certified Abstract Interpretation with Pretty-Big-Step Semantics*, in "Certified Programs and Proofs (CPP 2015)", Mumbai, India, Proceedings of the 2015 Conference on Certified Programs and Proofs, January 2015 [*DOI :* 10.1145/2676724.2693174], https://hal.inria.fr/hal-01111588.

[33] B. BOISSINOT, A. DARTE, F. RASTELLO, B. DUPONT DE DINECHIN, C. GUILLON.*Revisiting Out-of-SSA Translation for Correctness, Code Quality and Efficiency*, in "Proc. of CGO'09", IEEE Computer Society, 2009, p. 114–125.

[34] G. BOUDOL, G. PETRI, B. P. SERPETTE.*Relaxed Operational Semantics of Concurrent Programming Languages*, in "EXPRESS/SOS", 2012, vol. 89, p. 19-33.

[35] D. DEMANGE, V. LAPORTE, L. ZHAO, D. PICHARDIE, S. JAGANNATHAN, J. VITEK.*Plan B: A Buffered Memory Model for Java*, in "Proc. of the 40th ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages, POPL 2013", Rome, Italy, ACM, 2013, http://hal.inria.fr/hal-00924716.

[36] D. DEMANGE, L. STEFANESCO, D. PICHARDIE.*Verifying Fast and Sparse SSA-based Optimizations in Coq*, in "Proc. of CC'15", LNCS, 2015, vol. 9031, p. 233-252.

[37] J.-H. JOURDAN, V. LAPORTE, S. BLAZY, X. LEROY, D. PICHARDIE.*A formally-verified C static analyzer*, in "POPL 2015: 42nd ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages", Mumbai, India, ACM, January 2015, p. 247-259 [*DOI :* 10.1145/2676726.2676966], https://hal.inria.fr/hal-01078386.

[38] J. A. KROLL, G. STEWART, A. W. APPEL.*Portable Software Fault Isolation*, in "CSF 2014", IEEE, 2014, p. 18–32, http://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=6954678.

[39] P. WILKE.*Formally verified compilation of low-level C code*, University of Rennes 1, 2016.

Project-Team CIDRE

Confidentialité, Intégrité, Disponibilité et Répartition

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

IN PARTNERSHIP WITH:
CNRS

CentraleSupélec

Université Rennes 1

RESEARCH CENTER
Rennes - Bretagne-Atlantique

THEME
Distributed Systems and middleware

Table of contents

Project-Team CIDRE

Creation of the Project-Team: 2011 July 01

Keywords:

Computer Science and Digital Science:

1.2.8. - Network security
1.3. - Distributed Systems
3.3.1. - On-line analytical processing
3.5.2. - Recommendation systems
4.1.1. - Malware analysis
4.1.2. - Hardware attacks
4.4. - Security of equipment and software
4.8. - Privacy-enhancing technologies
4.9.1. - Intrusion detection
4.9.2. - Alert correlation
7.1. - Parallel and distributed algorithms

Other Research Topics and Application Domains:

6.5. - Information systems
9.8. - Privacy

1. Members

Research Scientists

Emmanuelle Anceaume [CNRS, Researcher]
Michel Hurfin [Inria, Researcher, HDR]

Faculty Members

Christophe Bidan [Team Leader, CentraleSupelec, Professor, HDR]
Sébastien Gambs [Univ. Rennes I, Associate Professor, until Jan 2016, HDR]
Gilles Guette [Univ. Rennes I, Associate Professor]
Guillaume Hiet [CentraleSupelec, Associate Professor]
Mohamed Kasraoui [Univ. Rennes I, Associate Professor]
Julien Lolive [Univ. Rennes I, Associate Professor, until Aug 2016]
Ludovic Mé [CentraleSupelec, Professor, HDR]
Guillaume Piolle [CentraleSupelec, Associate Professor]
Nicolas Prigent [CentraleSupelec, Associate Professor, until Sep 2016]
Eric Totel [CentraleSupelec, Professor, HDR]
Frédéric Tronel [CentraleSupelec, Associate Professor]
Valérie Viet Triem Tong [CentraleSupelec, Associate Professor, HDR]

Technical Staff

Christopher Humphries [Inria]

PhD Students

Solenn Brunet [Orange Labs, granted by CIFRE]
Damien Crémilleux [CentraleSupelec]
Aurélien Dupin [Thales, granted by CIFRE]
Laurent Georget [Univ. Rennes I]

Florian Grandhomme [Univ. Rennes I]
Antoine Guellier [Univ. Rennes I]
Kun He [Inst. de Recherche Technologique B-COM, until Sep 2016]
Mouna Hkimi [Inria, until Oct 2016]
David Lanoé [Inria, from Oct 2016]
Laetitia Leichtnam [Min. de la Défense, from Oct 2016]
Mourad Leslous [Inria]
Thomas Letan [CentraleSupelec]
Pernelle Mensah [Bell Labs (Alcatel), granted by CIFRE]
Mounir Nasr Allah [CentraleSupelec]
Deepak Subramanian [CentraleSupelec, until Aug 2016]
Aurélien Trulla [Inria, from Oct 2016]
Charles Arya Xosanavongsa [Thales, from Feb 2016, granted by CIFRE]

Post-Doctoral Fellow

Chuanyou Li [Inria, until Apr 2016]

Administrative Assistant

Lydie Mabil [Inria]

Others

Ronny Chevalier [CentraleSupelec, Internship, until Jul 2016]
Thibaut Lajoie-Mazenc [CNRS, Internship, from Jun 2016 until Nov 2016]
Frédéric Majorczyk [DGA, External Collaborator]
Evan Pisani [Univ. Rennes I, Internship, from May 2016 until Jul 2016]
Guereguin Der Sylvestre Sidibe [CentraleSupelec, Internship, from Apr 2016 until Aug 2016]
Jianqiao Xu [CentraleSupelec, Internship, from Jul 2016]

2. Overall Objectives

2.1. CIDRE in Brief

Our long term ambition is to contribute to the building of distributed systems that are trustworthy and respectful of privacy, even when some nodes in the system have been compromised.

With this objective in mind, the CIDRE team focuses mainly on the two following topics: Intrusion Detection and Privacy Protection.

3. Research Program

3.1. Our perspective

For many aspects of our everyday life, we heavily rely on information systems, many of which are based on massively networked devices that support a population of interacting and cooperating entities. While these information systems become increasingly open and complex, accidental and intentional failures get considerably more frequent and severe.

Two research communities traditionally address the concern of accidental and intentional failures: the distributed computing community and the security community. While both these communities are interested in the construction of systems that are correct and secure, an ideological gap and a lack of communication exist between them that is often explained by the incompatibility of the assumptions each of them traditionally makes. Furthermore, in terms of objectives, the distributed computing community has favored systems availability while the security community has focused on integrity and confidentiality, and more recently on privacy.

Our long term ambition is to contribute to the building of distributed systems that are trustworthy and respectful of privacy, even when some nodes [0] in the system have been compromised. For that purpose, we are convinced that combining classical security approaches and distributed computing paradigms is an interesting way to enforce the security of large-scale distributed systems. More specifically, since a distributed system is composed of nodes, we assert that the security of large-scale distributed systems has to be addressed at three complementary levels:

- the level of each node: each standalone node has to enforce its own security;
- the level of an *identified* set of *trusted* nodes: the *trusted* nodes can *collaborate* to enforce together their security;
- the level of fully open large-scale distributed and dynamic systems: distributed computing paradigms such as consensus algorithms can be applied to cope with the possible presence of malicious nodes.

Notice that using a distributed architecture can also be an approach allowing the nodes to enforce their security without the need of a trusted third party.

The research activities of the CIDRE project-team focus mainly on the two following research axis:

- **Intrusion Detection System:** the objective is to detect any suspicious events with regard to the security by analyzing some data generated on the monitored system.
- **Privacy-preserving Services:** the objective is to ensure users' privacy even when this property seems incompatible with the provided services, like social networks or location-based services.

In all our studies, we consider a priori that the attacker is omnipotent. He can acts as he wants. Nevertheless, being not a team specialized in cryptography, we consider that we can rely on strong unbroken cryto-systems.

3.2. Intrusion Detection / Security Events Monitoring and Management

Today, we are not yet fully entered into a world of "security by design". Security remains often a property that is considered a posteriori, when the system is deployed, which often results in applying patches when vulnerabilities are discovered (also called a "patch and pray" approach). Unfortunately, despite patching, the number of vulnerabilities remains high, as evidenced by the number of vulnerabilities published each year in the Common Vulnerabilities and Exposures (CVE) system. Thus, it is important to be able to early detect cyber-attacks, especially when they exploit vulnerabilities that are unknown. However, the efficiency of security events monitoring and management systems (including the IDS - Intrusion Detection Systems) is still an open issue today. Indeed, they are often unable to effectively deal with huge numbers of security events, and they usually produce too many false alarms yet missing some attacks. So one of the main research challenges in IT security remains the definition of efficient security events monitoring systems, i.e., that enable both to process a huge number of security events and to detect any attacks without flooding the security analysts with false alarms.

By exploiting vulnerabilities in operating systems, applications, or network services, an attacker can defeat preventive security mechanisms and violate the security policy of the whole system. The goal of an Intrusion Detection Systems (IDS) is to detect such violations by analyzing some *security events* generated on a monitored system. Ideally, the IDS should produce an alert for any violation (no *false negative*), and only for violations (no *false positive*).

To produce alerts, two detection techniques exist: the misuse based detection and the anomaly based detection. A misuse based detection is actually a signature based detection approach : it allows to detect only the attacks whose signature is available. From our point of view, while useful in practice, misuse detection is intrinsically limited. Indeed, it requires to update in real-time the database of signatures, similarly to what has to be done for antivirus tools. The CIDRE project-team follows the alternative approach, namely the anomaly approach, which consists in detecting a deviation from a referenced behavior. Our contributions on anomaly-based IDS follow three axis:

[0] The term node either refers to a device that hosts a network client or service or to the process that runs this client or service.

- **Illegal Information Flow Detection:** our goal is to detect information flows in the monitored system (either a node or a set of trusted nodes) that are allowed by the access control mechanism, but are illegal from the security policy point of view. This approach is particularly appealing to detect intrusions in a standalone node, such as a smartphone.
- **Anomaly-Based Detection in Distributed Applications:** our goal is to specify the normal behavior based on either a formal specification of the distributed application, or previous executions. This approach is particularly appealing to detect intrusions in industrial control systems since these systems exhibit well-defined behaviors at different levels: network level (network communication patterns, protocol specifications, etc.), control level (continue and discrete process control laws), or even the state of the local resources (memory or CPU).
- **Online data analytics:** our goal is to estimate on the fly different statistics or metrics on distributed input streams to detect abnormal behavior with respect to a well-defined criterion such as the distance between different streams, their correlation or their entropy.

Beside the anomaly-based IDS, we have also led research work on alert correlation and visualisation of security events. Indeed, in large systems, multiple (host and network) IDS and many sensors are deployed and they continuously and independently generate notifications (event's observations, warnings and alerts). To cope with this huge amount of collected data, we have studied two different approaches, each with specific goal:

- **Alert Correlation System:** the alerts of *low level* IDSes can be viewed as *security events* of a *high level* IDS whose goal is to correlate these alerts. An alert correlation system aims at exploiting the known relationships between some elements that appear in the flow of low level notifications to generate high semantic meta-alerts. The main goal is to reduce the number of alerts (and especially, false positive) returned to the security analysts and to allow a higher level analysis of the situation (situational awareness).
- **Visualization Tools:** a visualization tools aims at relying on the capacity of human beings to detect patterns and outliers in datasets when these datasets are properly visually represented. Human beings also know pieces of contextual information that are very difficult to formalize so as to make them usable by a computer. Visualization is therefore a very useful complementary tool to detect abnormal events in real time (monitoring), to search for malicious events in log files (data exploration and forensics) and to communicate results (reporting).

3.3. Privacy

In a world of ubiquitous technologies, each individual constantly leaves digital traces related to his activities and interests. The current business plan of many web services such as social networks, is based on the sale of these digital traces. Of course, this is usually done in a legal way, the license of use clearly stating that the user gives the right to the service provider for using his personal data. However, on the one hand, users generally do not read these licenses, and on the other hand, these licenses are usually very vague on the use of personal data [0]. In addition these digital traces can potentially be stolen and maliciously used, they must therefore be protected. In this context, users' privacy is now recognized as a fundamental individual right. Any new IT service should thus follow the *privacy-by-design* approach: privacy issues have to be studied from the earliest phase of a project by taking into account the multi-stakeholders and transdisciplinary aspects in order to ensure proper, end-to-end private data protection properties.

In the CIDRE project, we mainly focus on domains in which privacy issues collide with provided services. Here are some concrete examples of such domains:

- **Location-based services:** the challenge is to design services that depend on the user's location while preserving the privacy of his location;
- **Social networks:** the challenge is to demonstrate that it is possible to design social networks respectful of users' privacy;

[0] Besides, it has been shown that service providers do not necessarily comply with their own license.

- **Mobile services:** given that such services are based on user's identity, the challenge is to design mobile services while preserving the users' anonymity;
- **Ad-hoc netwoks:** in ad-hoc networks, any participant can potentially know the relative location of the other participants. Thus, the issue is to allow nodes to forward messages while preserving the privacy of the communications.

For all of these domains, we have proposed new Privacy-Enhancing Techniques (PETs) based on a mix of different foundations such as cryptographic techniques, security policies and access control mechanisms, just to name a few. More generally, we think that a major option to protect users' privacy consists in using a decentralized architecture that enables to transfer control and services from the service providers to the users.

The concept of IDS seems to be in contradiction with the users' privacy. Indeed, an IDS is a monitoring system that needs to collect and analyze information coming from different levels such as network, applications and OS, this information being able to include users' personal data. However, we are confident that IDS and privacy are not completely antagonist. In particular, integrating some privacy features inside an IDS to build a privacy-preserving IDS may allow to limit the amount of information that can leak if one of the nodes within the system is compromised. On the other hand, enabling IDS to detect attacks against privacy as well as security violations can extend the range of their applicability.

4. Application Domains

4.1. Security is Required Everywhere

With the infiltration of computers and software in almost all aspects of our modern life, security can nowadays be seen as an absolutely general concern. As such, the results of the research targeted by CIDRE apply to a wide range of domains. It is clear that critical systems, in which security (and safety) is a major concern can benefit from ideas such as dynamic security policy monitoring. On the other hand, systems used by the general public (basically, the internet and services such as web or cloud services, social networks, location-based services, etc.) can also benefit from results obtained by CIDRE, in particular to solve some of the privacy issues raised by these systems that manipulate huge amount of personal data. In addition, systems are getting more and more complex, decentralized, distributed, or spontaneous. Cloud computing, in particular, brings many challenges that could benefit from ideas, approaches and solutions studied by CIDRE in the context of distributed systems.

Industrial Control Systems (ICS) and in particular Supervisory Control and Data Acquisition are also new application domains for intrusion detection. The Stuxnet attack has emphasized the vulnerability of such critical systems which are not totally isolated anymore. Securing ICS is challenging since modifications of the systems, for example to patch them, are often not possible. High availability requirements also often conflict with preventive approaches. In this case, security monitoring is appealing to protect such systems against malicious activities. Intrusion detection in ICS is not fundamentally different from traditional approaches. However, new hypotheses and constraints need to be taken into account, which also bring interesting new research challenges.

5. Highlights of the Year

5.1. Highlights of the Year

5.1.1. Awards

Mounir Assaf, a former PhD student, has received the "prix de thèse du GDR GPL" in June 2016. His PhD thesis is entitled "évaluation des fuites d'information dans les logiciels critiques" and has been defended in 2015.

Emmanuelle Anceaume has received the Most Prolific Author Award during the NCA conference.
BEST PAPERS AWARDS :
[28] **9th International Conference on Security of Information and Networks (SIN 2016)**. D. SUBRAMANIAN, G. HIET, C. BIDAN.
[24] **Symposium on Network Computing and Applications**. Y. MOCQUARD, B. SERICOLA, S. ROBERT, E. ANCEAUME.

6. New Software and Platforms

6.1. Blare

To detect intrusion using information flows
KEYWORDS: Cybersecurity - Intrusion Detection Systems (IDS) - Data Leakage Protection
SCIENTIFIC DESCRIPTION

Blare implements our approach of illegal information flow detection for a single node (Android and Linux kernel, JVM) and a set of nodes (monitoring of flows between linux machines).
FUNCTIONAL DESCRIPTION

Blare IDS is a set of tools that implements our approach to illegal information flow detection for a single node and a set of nodes.

- Partner: SUPELEC
- Contact: Frédéric Tronel
- URL: http://blare-ids.org

6.2. ELVIS

Extensible Log VISualization
KEYWORDS: Visualization - Cybersecurity - Intrusion Detection Systems (IDS) - SIEM - Cyber attack - Forensics
SCIENTIFIC DESCRIPTION

The studies that were performed since 2012 clearly showed that there was an important need for technologies that would allow analysts to handle in a consistent way the various types of log files that they have to study in order to detect intrusion or to perform forensic analysis. Consequently, we proposed this year ELVis, a security-oriented log visualization system that allows the analyst to import its log files and to obtain automatically a relevant representation of their content based on the type of the fields they are made of. First, a summary view is proposed. This summary displays in an adequate manner each field according to its type (i.e. categorical, ordinal, geographical, etc.). Then, the analyst can select one or more fields to obtain some details about it. A relevant representation is then automatically selected by the tool according to the types of the fields that were selected.

ELVis [35] has been presented in VizSec 2013 (part of Vis 2013) in October 2013 in Atlanta. A working prototype is currently being tuned in order to perform field trials with our partners in DGA-MI. Next year, we are planing to perform research on how various log files can be combined in the same representation.
FUNCTIONAL DESCRIPTION

ELVIS is a visualisation tool geared to system security which enables analysts to visually explore log files using relevant representations. The tool accepts many different types of log file and can easily be extended to accept new ones opportunistically. Thanks to its data typing mechanisms, it can automatically choose relevant representations depending on the type of data that the analyst wants to observe.

- Participant: Nicolas Prigent
- Partner: SUPELEC
- Contact: Nicolas Prigent
- URL: https://hal.inria.fr/hal-00875668

6.3. GEPETO

GEoPrivacy-Enhancing TOolkit
KEYWORDS: Cyber attack - Privacy - Mobility
SCIENTIFIC DESCRIPTION

(GEoPrivacy-Enhancing TOolkit) is an open source software for managing location data (currently in development in cooperation with LAAS). GEPETO can be used to visualize, sanitize, perform inference attacks and measure the utility of a particular geolocated dataset. For each of these actions, a set of different techniques and algorithms can be applied. The global objective of GEPETO is to enable a user to design, tune, experiment and evaluate various sanitization algorithms and inference attacks as well as visualizing the following results and evaluating the resulting trade-off between privacy and utility. An engineer (Izabela Moïse) has contributed to the development of a distributed version of GEPETO based on the MapReduce paradigm and the Hadoop framework that is able to analyze datasets composed of millions of mobility traces in a few minutes [30] .
FUNCTIONAL DESCRIPTION

GEPETO is an open source software for managing location data. GEPETO can be used to visualize, sanitize, perform inference attacks, and measures the utility of a particular geolocated dataset.

- Partners: CNRS - Université de Rennes 1
- Contact: Sébastien Gambs
- URL: https://gforge.inria.fr/projects/gepeto/

6.4. GNG

Security Supervision by Alert Correlation
KEYWORDS: Intrusion Detection Systems (IDS) - SIEM
SCIENTIFIC DESCRIPTION

GNG is an intrusion detection system that correlates different sources (such as different logs) in order to identify attacks against the system. The attack scenarios are defined using the Attack Description Langage (ADeLe) proposed by our team, and are internally translated to attack recognition automatons. GNG intends to define time efficient algorithms based on these automatons to recognize complex attack scenarios.

- Partner: SUPELEC
- Contact: Eric Totel
- URL: http://www.rennes.supelec.fr/ren/perso/etotel/GNG/index.html

6.5. GroddDroid

KEYWORDS: Android - Detection - Malware
FUNCTIONAL DESCRIPTION

GroddDroid

1- locates suspicious code in Android application

2-computes execution paths towards suspicious code

3- forces executions of suspicious code

- Partners: CentraleSupélec - Insa Centre Val-de-Loire
- Contact: Valérie Viet Triem Tong
- URL: http://kharon.gforge.inria.fr/grodddroid.html

6.6. Kharon platform

KEYWORDS: Android - Malware - Dynamic Analysis
FUNCTIONAL DESCRIPTION

This platform executes Android applications and computes a graph representing all the information flows that occurred in the operating system due to a malicious execution. It can then classify observed behavior as benign or malicious. Access to this platform is currently in physically controlled at the high security laboratory (LHS) of Rennes.

- Partners: CentraleSupélec - Insa Centre Val-de-Loire
- Contact: Valérie Viet Triem Tong
- URL: http://kharon.gforge.inria.fr/

6.7. Netzob

FUNCTIONAL DESCRIPTION

Netzob is an opensource tool for reverse engineering, traffic generation and fuzzing of communication protocols. This tool allows to infer the message format (vocabulary) and the state machine (grammar) of a protocol through passive and active processes. Its objective is to bring state of art academic researches to the operational field, by leveraging bio-informatic and grammatical inferring algorithms in a semi-automatic manner.

- Participant: Georges Bossert
- Contact: Ludovic Mé
- URL: http://www.netzob.org/

6.8. VEGAS

Visualizing, Exploring and Grouping AlertS
KEYWORDS: Security - Visualization - Cybersecurity - Intrusion Detection Systems (IDS) - SIEM
SCIENTIFIC DESCRIPTION

VEGAS explore the hypothesis that is is possible to offer to front-line security operators a visualization tool that allows the to perform a first informed triage of the alerts that were received from IDSes so as to group them and transmit them to security analysts in a relevant way.
FUNCTIONAL DESCRIPTION

VEGAS is a visualization tool that allows to easily identify, explore and group alerts generated by an IDS. This tool allows security operators to easily dispatch similar alerts to security analyst to help them study them more efficiently.

- Participants: Damien Cremilleux, Frédéric Majorczyk and Nicolas Prigent
- Partner: SUPELEC
- Contact: Damien Crémilleux

7. New Results

7.1. Intrusion Detection

7.1.1. Intrusion Detection in Distributed Systems

Alert Correlation: In large systems, multiple (host and network) Intrusion Detection Systems (IDS) and many sensors are usually deployed. They continuously and independently generate notifications (event's observations, warnings and alerts). To cope with this amount of collected data, alert correlation systems have to be designed. An alert correlation system aims at exploiting the known relationships between some elements that appear in the flow of low level notifications to generate high semantic meta-alerts. The main goal is to reduce the number of alerts returned to the security administrator and to allow a higher level analysis of the situation. However, producing correlation rules is a highly difficult operation, as it requires both the knowledge of an attacker, and the knowledge of the functionalities of all IDSes involved in the detection process. In the context of the PhD of Erwan Godefroy [1], we focus on the transformation process that allows to translate the description of a complex attack scenario into correlation rules and its assessment. We show that, once a human expert has provided an action tree derived from an attack tree, a fully automated transformation process can generate exhaustive correlation rules that would be tedious and error prone to enumerate by hand.

Long lived attack campaigns known as Advanced Persistent Threats (APTs) have emerged as a serious security risk. These attack campaigns are customised for their target and performed step by step during months on end. The major difficulty in detecting an APT is keeping track of the different steps logged over months of monitoring and linking them. In [11], we describe TerminAPTor, an APT detector which highlights links between the traces left by attackers in the monitored system during the different stages of an attack campaign. TerminAPTor tackles this challenge by resorting to Information Flow Tracking (IFT). Our main contribution is showing that IFT can be used to highlight APTs. Additionally, we describe a generic representation of APTs and validate our IFT-based APT detector.

Inferring the normal behavior of an application: In [29], [6], [41], we propose an approach to detect intrusions that affect the behavior of distributed applications. To determine whether an observed behavior is normal or not (occurence of an attack), we rely on a model of normal behavior. This model has been built during an initial training phase (machine learning approach). During this preliminary phase, the application is executed several times in a safe environment. The gathered traces (sequences of actions) are used to generate an automaton that characterizes all these acceptable behaviors. To reduce the size of the automaton and to be able to accept more general behaviors that are close to the observed traces, the automaton is transformed. These transformations may lead to introduce unacceptable behaviors. Our current work aims at identifying the possible errors tolerated by the compacted automaton.

This approach is particularly appealing to detect intrusions in industrial control systems since these systems exhibit well-defined behaviors at different levels: network level (network communication patterns, protocol specifications, etc.), control level (continue and discrete process control laws), or even the state of the local resources (memory or CPU). Industrial control systems (ICS) can be subject to highly sophisticated attacks which may lead the process towards critical states. Due to the particular context of ICS, protection mechanisms are not always practical, nor sufficient. On the other hand, developing a process-aware intrusion detection solution with satisfactory alert characterization remains an open problem. In [20], we focus on process-aware attacks detection in sequential control systems. We build on results from runtime verification and specification mining to automatically infer and monitor process specifications. Such specifications are represented by sets of temporal safety properties over states and events corresponding to sensors and actuators. The properties are then synthesized as monitors which report violations on execution traces. We develop an efficient specification mining algorithm and use filtering rules to handle the large number of mined properties. Furthermore, we introduce the notion of activity and discuss its relevance to both specification mining and attack detection in the context of sequential control systems. The proposed approach is evaluated in a hardware-in-the-loop setting subject to targeted process-aware attacks. Overall, due to the explicit handling of process variables, the solution provides a better characterization of the alerts and a more meaningful understanding of false positives.

7.1.2. Illegal Information Flow Detection

Our research work on intrusion detection based on information flow has been initiated in 2002. This research work has resulted in Blare, a framework for Intrusion Detection Systems [0], including KBlare, an implementation as a Linux Security Module (LSM), JBlare, an implementation for the Java Virtual Machine (JVM), and AndroBlare, for Android applications.

Illegal Information Flow in Web-browser: In the context of the CominLabs SECCLOUD project, we were interested in implementing our approach to detect illegal information flow in web-browser. We have proposed a new secure information flow control model specifically designed for JavaScript [28]. In our approach, we augment the standard symbol table with a mechanism that replaces the reference address for secret values based on the current execution stack. This mechanism also ensures that the secret is stored in a dedicated memory location thereby protecting the secret from any unintended leakage or modification by a malicious JavaScript. This work on detection of illegal information flow in JavaScript has received the best paper award at the 9th International Conference on Security of Information and Networks (SIN 2016) [28].

Later Deepak Subramanian has improved this approach and optimized the computation time required to determine the legacy of information flows. An approach which begins with a learning phase allows to increase

[0] http://www.blare-ids.org

the accuracy of the proposed solution. Information about the modified variables are kept in memory to perform a more accurate analysis of the indirect information flows. This self-correcting information flow control model for a web-browser is described in [27].

Information Leaks: Qualitative information flow aims at detecting information leaks, whereas the emerging quantitative techniques target the estimation of information leaks. Quantifying information flow in the presence of low inputs is challenging, since the traditional techniques of approximating and counting the reachable states of a program no longer suffice. In [32], we propose an automated quantitative information flow analysis for imperative deterministic programs with low inputs. The approach relies on a novel abstract domain, the cardinal abstraction, in order to compute a precise upper-bound over the maximum leakage of batch-job programs. We prove the soundness of the cardinal abstract domain by relying on the framework of abstract interpretation. We also prove its precision with respect to a flow-sensitive type system for the two-point security lattice.

More generally, for his research activities during his PhD thesis, Mounir Assaf has received the 2016 thesis prize awarded by the GDR GPL (Engineering Programming and Software).

Characterizing Android Malwares: Android has become the world's most popular mobile operating system, and consequently the most popular target for unscrupulous developers. These developers seek to make money by taking advantage of Android users who customise their devices with various applications, which are the main malware infection vector. Indeed, the most likely way a user executes a repackaged application is by downloading a seemingly harmless application from a store and executing it. Such an application may have been modified by an attacker in order to add malicious pieces of code.

To fight repackaged applications containing malicious code, most official application marketplaces have implemented security analysis tools that try to detect and remove malware. Countermeasures adopted by the attackers to bypass these new controls can be divided into two main approaches: avoiding static analysis and avoiding dynamic analysis [39]. A static analysis of an application consists of analysing its code and its resources without executing it. Conversely, dynamic analysis stands for any kind of analysis that requires executing the application in order to observe its actions.

The Kharon project [19] goes a step further from classical dynamic analysis of malware (http://kharon.gforge.inria.fr). Funded by the Labex CominLabs and involving partners of Centrale-Supélec, Inria and INSA Centre Val de Loire, this project aims to capture a compact and comprehensive representation of malware. To achieve such a goal we have developed tools to monitor operating systems' information flows induced by the execution of a marked application. We support the idea that the best way to understand malware impact is to observe it in its normal execution environment i.e., a real smartphone. Additionally, the main challenge is to be able to trigger malicious behaviours even if the malware tries to escape dynamic analysis.

In this context, we have developed an original solution that mainly consists of 'helping the malware to execute'. In other words we slightly modify the bytecode of the infected application in order to defeat the protection against dynamic analysis and we execute the suspicious code in its most favourable execution conditions. Thus, our software helps us understand malware's objectives and the consequences on the health of a user's device. In particular, we use a global control flow graph (CFG) to exhibit an execution path to reach specific parts of code [42].

To achieve stealthiness when attacking a mobile device, an effective approach is the use of a covert channel built by two colluding applications to locally exchange data. Since this process is tightly coupled with the used hiding method, its detection is a challenging task, also worsened by the very low transmission rates. Using general indicators such as the energy consumed by the device, we propose in [5] an approach to detect the hidden data exchange between colluding applications and show its feasibility and effectiveness through different experimental results.

Our main research direction and challenge is to develop new and original protections against malicious applications that try to defeat classical dynamic analysis.

7.1.3. Intrusion Detection in Low-Level Software Components

In order to protect the IDS itself, we have initiated different research activities in the domain of hardware security. Our goal is to use co-design software/hardware approaches against traditional software attacks. In a bilateral research project with HP Inc Research Labs, we investigate how dedicated hardware could be used to monitor the whole software stack (from the firmware to the user-mode applications). In the CominLabs HardBlare project, we study the use of a dedicated co-processor to enforce Dynamic Information Flow Control on the main CPU. Finally, in the context of the PhD thesis of Thomas Lethan (ANSSI), we investigate the use of formal methods to evaluate the security guarantees provided by hardware platforms, which combine different CPUs, chipsets and memories. Over time, hardware designs have constantly grown in complexity and modern platforms involve multiple interconnected hardware components. During the last decade, several vulnerability disclosures have proven that trust in hardware can be misplaced. In [21], [37], we give a formal definition of Hardware-based Security Enforcement (HSE) mechanisms, a class of security enforcement mechanisms such that a software component relies on the underlying hardware platform to enforce a security policy. We then model a subset of a x86-based hardware platform specifications and we prove the soundness of a realistic HSE mechanism within this model using Coq, a proof assistant system.

The HardBlare project proposes a software/hardware co-design methodology to ensure that security properties are preserved all along the execution of the system but also during files storage. It is based on the Dynamic Information Flow Tracking (DIFT) that generally consists in attaching tags to denote the type of information that are saved or generated within the system. These tags are then propagated when the system evolves and information flow control is performed in order to guarantee the safe execution and storage within the system monitored by security policies [43].

In [30] we introduce an efficient approach for DIFT (Dynamic Information Flow Tracking) implementations on reconfigurable chips. Existing solutions are either hardly portable or bring unsatisfactory time overheads. This work presents an innovative implementation for DIFT on reconfigurable SoCs such as Xilinx Zynq devices.

In [7], we detail a hardware-assisted approach for information flow tracking implemented on reconfigurable chips. Current solutions are either time-consuming or hardly portable (modifications of both software/hardware layers). This work takes benefits from debug components included in ARMv7 processors to retrieve details on instructions committed by the CPU. First results in terms of silicon area and time overheads are also given.

7.1.4. Vizualization

The large quantities of alerts generated by intrusion detection systems (IDS) make very difficult to distinguish on a network real threats from noise. To help solving this problem, we propose VEGAS [12], an alerts visualization and classification tool that allows first line security operators to group alerts visually based on their principal component analysis (PCA) representation. VEGAS is included in a workflow in such a way that once a set of similar alerts has been collected and diagnosed, a filter is generated that redirects forthcoming similar alerts to other security analysts that are specifically in charge of this set of alerts, in effect reducing the flow of raw undiagnosed alerts.

Our research on visualization of security events has lead to two proofs-of-concept (See ELVIS and VEGAS softwares). We are currently pursuing business opportunities on this topic. Indeed SplitSec is a soon to be founded startup developing tools to help security experts to better manage and understand security data. Scalable analysis solutions and data visualisations adapted for security are combined into powerful tools for incident response. Christopher Humphries is a technology transfer engineer employed by Inria to build these tools based on promising research prototypes.

7.2. Privacy

7.2.1. Image Encryption

More and more users prefer to share their photos through image-sharing platforms of social networks than using e-mail or personal webpages. Since the provider of the image-sharing platform can clearly know the contents of any published images, the users have to trust the provider to respect their privacy or has to encrypt their images. In the context of the PhD of Kun He [18], [17], [16], we have proposed an IND-CPA image encryption algorithm that preserve the image format after encryption, and we have shown that our encryption algorithm can be used on several widely used image-sharing platforms such as Flickr, Pinterest, Google+ and Twitter.

7.2.2. Fingerprinting

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. In the context of the PhD of Julien Lolive, we have defined the first privacy-preserving asymmetric fingerprinting protocol based on Tardos codes [2]. This protocol is optimal with respect to traitor tracing. We also formally proved that our protocol achieves the properties of correctness, anti-framing, traitor tracing, as well as buyer- and item-unlinkability.

7.3. Communication and Synchronization in Distributed Systems

7.3.1. Routing Protocol for Tactical Mobile Ad Hoc Networks

In the context of the PhD thesis of Florian Grandhomme, we propose new secure and efficient algorithms and protocols to provide inter-domain routing in the context of tactical mobile ad hoc network. The proposed protocol has to handle context modification due to the mobility of Mobile Ad hoc NETwork (MANET), that is to say split of a MANET, merge of two or more MANET, and also handle heterogeneity of technology and infrastructure. The solution has to be independent from the underlying intra-domain routing protocol and from the infrastructure: wired or wireless, fixed or mobile. This work is done in cooperation with DGA-MI.

New generation military equipment, soldiers and vehicles, use wireless technology to communicate on the battlefield. During missions, they form a MANET. Since the battlefield includes coalition, each group may communicate with another group, and inter-MANET communication may be established. Inter-MANET (or inter-domain MANET) communication should allow communication, but maintain a control on the exchanged information. Several protocols have been proposed in order to handle inter-domain routing for tactical MANETs. In [14], [33], we describe and compare three solutions. Based on this analysis, we propose some preconizations to design Inter-domain protocols for MANET.

In [15], we present a coalition context and describe the functional hypothesis we used. Then, we propose a protocol that would fit such a network and conduct experimentation that tend to show that our proposition is quite efficient.

7.3.2. Communication and Synchronization Primitives

Use of Primitives to Limit Equivocation: We consider the approximate consensus problem in a partially connected network of n nodes where at most f nodes may suffer from Byzantine faults. In [22], we study under which conditions this problem can be solved using an iterative algorithm. A Byzantine node can equivocate: it may provide different values to its neighbors. To restrict the possibilities of equivocation, the 3-partial multicast primitive is considered. When a (correct or faulty) node uses this communication primitive, it provides necessarily the same value to the two identified receivers. Based on this communication primitive, a novel condition called f-resilient is proposed and proved to be necessary and sufficient to solve the approximate Byzantine consensus problem in a synchronous network.

The Test&Set Problem: In [35], we present a solution to the well-known problem of synchronization in a distributed asynchronous system prone to process crashes. This problem is also known as the Test&Set problem. The Test&Set is a distributed synchronization protocol that, when invoked by a set of processes, returns a unique winning process. This unique process is then allowed to use, for instance, a shared resource. Recently many advances in implementing Test&Set objects have been achieved, however all of them uniquely target the shared memory model. In this paper we propose an implementation of a Test&Set object for a message passing distributed system. This implementation can be invoked by any number $n \leq N$ of processes where N is the total number of processes in the system. We show in this paper, using a Markov model, that our implementation has an expected step complexity in O(log n) and we give an explicit formula for the distribution of the number of steps needed to solve the problem.

7.3.3. Dependability in Cloud Storage

The quantity of data in the world is steadily increasing bringing challenges to storage system providers to find ways to handle data efficiently in terms of dependability and in a cost-effectively manner. We have been interested in cloud storage which is a growing trend in data storage solution. For instance, the International Data Corporation (IDC) predicts that by 2020, nearly 40% of the data in the world will be stored or processed in a cloud. The thesis of Pierre Obame [3] addressed challenges around data access latency and dependability in cloud storage. We proposed Mistore, a distributed storage system that we designed to ensure data availability, durability, low access latency by leveraging the Digital Subscriber Line (xDSL) infrastructure of an Internet Service Provider (ISP). Mistore uses the available storage resources of a large number of home gateways, Points of Presence, and datacenters for content storage and caching facilities. Mistore also targets data consistency by providing multiple types of data consistency criteria and a versioning system. We also considered the data security and confidentiality in the context of storage systems applying data deduplication which is becoming one of the most popular data technologies to reduce the storage cost and we design a data deduplication method that is secure against malicious clients while remaining efficient in terms of network bandwidth and storage space savings.

7.3.4. Decentralized Cryptocurrency Systems

Decentralized cryptocurrency systems offer a medium of exchange secured by cryptography, without the need of a centralized banking authority. Among others, Bitcoin is considered as the most mature one [10]. Its popularity lies on the introduction of the concept of the blockchain, a public distributed ledger shared by all participants of the system. Double spending attacks and blockchain forks are two main issues in blockchain-based protocols. The first one refers to the ability of an adversary to use the very same bitcoin more than once, while blockchain forks cause transient inconsistencies in the blockchain. In [9], we show through probabilistic analysis that the reliability of recent solutions that exclusively rely on a particular type of Bitcoin actors, called miners, to guarantee the consistency of Bitcoin operations, drastically decreases with the size of the blockchain.

Some recent works have proposed to improve upon Bitcoin weaknesses. In [31], we analyze of one of these recent works, and shows through an analytical performance evaluation that new Bitcoin improvements are still needed.

7.3.5. Large Scale Systems

Population Protocol: the computational model of population protocols is a formalism that allows the analysis of properties emerging from simple and pairwise interactions among a very large number of anonymous finite-state agents. Significant work has been done so far to determine which problems are solvable in this model and at which cost in terms of states used by the protocols and time needed to converge. The problem tackled in [23] is the population proportion problem: each agent starts independently from each other in one of two states, say A or B, and the objective is for each agent to determine the proportion of agents that initially started in state A, assuming that each agent only uses a finite set of state, and does not know the number n of agents. We propose a solution which guarantees with any high probability that after O(log n) interactions any agent outputs with a precision given in advance, the proportion of agents that start in state A. The population proportion problem is a generalization of both the majority and counting problems, and thus our solution solves both problems. We show that our solution is optimal in time and space. Simulation results illustrate our theoretical analysis.

Propagation Time of a Rumor: the context of this work is the well studied dissemination of information in large scale distributed networks through pairwise interactions. This problem, originally called rumor mongering, and then rumor spreading has mainly been investigated in the synchronous model. This model relies on the assumption that all the nodes of the network act in synchrony, that is, at each round of the protocol, each node is allowed to contact a random neighbor. In [24], we drop this assumption under the argument that it is not realistic in large scale systems. We thus consider the asynchronous variant, where at time unit, a single node interacts with a randomly chosen neighbor. We perform a thorough study of the total number of interactions needed for all the nodes of the network to discover the rumor.

Distributed Stream Processing Systems: shuffle grouping is a technique used by stream processing frameworks to share input load among parallel instances of stateless operators. With shuffle grouping each tuple of a stream can be assigned to any available operator instance, independently from any previous assignment. A common approach to implement shuffle grouping is to adopt a Round-Robin policy, a simple solution that fares well as long as the tuple execution time is almost the same for all the tuples. However, such an assumption rarely holds in real cases where execution time strongly depends on tuple content. As a consequence, parallel stateless operators within stream processing applications may experience unpredictable unbalance that, in the end, causes undesirable increase in tuple completion times. In [25], [26] we propose Online Shuffle Grouping (OSG), a novel approach to shuffle grouping aimed at reducing the overall tuple completion time. OSG estimates the execution time of each tuple, enabling a proactive and online scheduling of input load to the target operator instances. Sketches are used to efficiently store the otherwise large amount of information required to schedule incoming load. We provide a probabilistic analysis and illustrate, through both simulations and a running prototype, its impact on stream processing applications.

Load shedding is a technique employed by stream processing systems to handle unpredictable spikes in the input load whenever available computing resources are not adequately provisioned. A load shedder drops tuples to keep the input load below a critical threshold and thus avoid unbounded queuing and system trashing. In [38] we propose Load-Aware Shedding (LAS), a novel load shedding solution that, unlike previous works, does not rely neither on a pre-defined cost model nor on any assumption on the tuple execution duration. Leveraging sketches, LAS efficiently builds and maintains at runtime a cost model to estimate the execution duration of each tuple with small error bounds. This estimation enables a proactive load shedding of the input stream at any operator that aims at limiting queuing latencies while dropping as few tuples as possible. We provide a theoretical analysis. Furthermore, through an extensive practical evaluation based on simulations and a prototype, we evaluate its impact on stream processing applications, which validate the robustness and accuracy of LAS.

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Contracts with Industry

- **HP (2013-2016): Embedded Systems Security**

  We aim at researching and prototyping low-level intrusion detection mechanisms in embedded system software. This involves mechanisms in continuation of previous work realized by our team as well as investigating new techniques more directly tied to specific HP device architectures. Our main objective is to monitor low-level software (firmware, OS kernels, hypervisors) thanks to a dedicated external co-processor. Being under NDA, details about this research program cannot be provided.

8.2. Bilateral Grants with Industry

- **Orange Labs: Privacy-preserving location-based services**

  Solenn Brunet has started her PhD thesis in September 2014 within the context of a CIFRE contract with Orange Labs Caen. Her PhD subject concerns the development of privacy-preserving location-based services that are able to personalize the service provided to the user according to his current position while preserving his location privacy. In particular, Solenn Brunet adapts existing cryptographic primitives (private information retrieval, secure multiparty computation, secure set intersection, ...) or design novel ones to use them as building blocks for the construction of these privacy-preserving location-based services.

- **DGA: BGP-like Inter Domain routing protocol for tactical mobile ad hoc networks: feasibility, performances and quality of service**

  Florian Grandhomme has started his PhD thesis in October 2014 in cooperation with DGA-MI. The subject of the PhD is to propose new secure and efficient algorithms and protocols to provide inter-domain routing in the context of tactical mobile ad hoc network. The protocol proposed will have to handle context modification due to the mobility of MANET, that is to say split of a MANET, merge of two or more MANET, and also handle heterogeneity of technology and infrastructure. The solution will have to be independent from the underlying intra-domain routing protocol and from the infrastructure: wired or wireles, fixed or mobile.

- **DGA: Visualization for security events monitoring**

  Damien Crémilleux has started his PhD thesis in October 2015 in the context of a cooperation with DGA-MI. The subject of the PhD is to define relevant representations to allow front-line security operators to monitors systems from a security perspective. A first proposal was made that led to a tool, VEGAS, that allows to monitor large quantities of alerts in real time and to dispatch these alerts in a relevant way to security analysts.

- **DGA: Intrusion Detection in Distributed Applications**

  David Lanoé has started his PhD thesis in October 2016 in the context of a cooperation with DGA-MI. His work will focus on the construction of behavioral models (during a learning phase) and their use to detect intrusions during an execution of the modelled distributed application.

- **Nokia: Risk-aware security policies adaptation in modern communication infrastructures**

  Pernelle Mensah was hired in January 2016 on this CIFRE funding in order to work on unexplored aspects of information security, and in particular response strategies to complex attacks, in the context of cloud computing architectures. The use case proposed by our industrial partner is a multi-tenant cloud computing platform involving software-defined networking in order to provide further flexibility and responsiveness in architecture management. The topic of the thesis is to adapt and improve the current risk-aware reactive response tools, based on attack graphs and adaptive security policies, to this specific environment, taking into account the heterogeneity of actors, platforms, policies and remediation options.

- **B-Com: Privacy Protection for JPEG Content on Image-Sharing Platforms**

  Kun He was hired as a PhD in September 2013 by the IRT B-Com. The subject of the PhD was the protection of users' privacy while publishing images on image-sharing platforms. The proposed solution is an image encryption algorithm that preserve the image format after encryption, and the experimentation have shown that the proposed encryption algorithm can be used on several widely used image-sharing platforms such as Flickr, Pinterest, Google+, Facebook and Twitter.

- **Thalès: Privacy and Secure Multi-party Computation**

  Aurélien Dupin has started his PhD thesis in January 2016 within the context of a CIFRE contract with Thalès. His PhD subject concerns the development of privacy-preserving location-based services based on secure multi-party computation. As part of his Master of Science from the ETS (Ecole de Technologie Supérieure) in Montreal, co-supervised by Prof. Jean-Marc ROBERT (ETS) and Prof. Christophe BIDAN (CentraleSupélec), Mr Aurélien DUPIN has already addressed the issue and proposed multi-party computation protocols to provide evidence of geolocations while ensuring the secrecy of the geographical location of participants protocols. The thesis is an opportunity to continue the work initiated during the Master of Science.

- **Thalès: Combining Attack Specification and Dynamic Learning from traces for correlation rule generation**

  Charles Xosanavongsa has started his PhD thesis in December 2016 in the context of a CIFRE with Thales. His work will focus on the construction of correlation rules. In previous work on correlation rule generation, the usual approach is static. It always relies on the description of the supervised system using a knowledge base of the system. The use of correlation trees is an appealing solution because it allows to have a precise description of the attacks and can handle any kind of IDS. But in practice, the behavior of each IDS is quite difficult to predict, in particular for anomaly based IDS. To manage automatically the correlation rules (and adapt them if necessary), we plan to analyze synthetic traces containing both anomaly based and misused based IDS alerts resulting from an attack.

9. Partnerships and Cooperations

9.1. Regional Initiatives

- **Region Bretagne ARED Grant :** the PhD of Mourad Leslous on malicious codes in Android applications is supported by a grant from the Région Bretagne.
- **Labex COMINLAB contract (2012-2016): "SecCloud" - http://www.seccloud.cominlabs.ueb.eu/** Attacks targeting web browsers constitute a major threat. We tackled in the context of the SecCloud project attacks induced by client-side code execution (javascript, flash or html5). Existing security mechanisms such as os-level access control often are not sufficient to pre- vent client-side browser attacks as the web browser is granted the same privileges as the user. The idea is to monitor information flows within the web browser in order to enforce a security information flow policy. Such a policy should allow to define fine-grained in- formation flow rules between user data and distant web sites. We proposed a new secure information flow control model specifically designed for JavaScript. This study was conducted in cooperation with other Inria Teams (Ascola and Celtique). Deepak Subramanian is doing his PhD in the context of this project.
- **Labex COMINLAB contract (2013-2018): "DeSceNt" - http://www.descent.cominlabs.ueb.eu** In DeSceNt, we propose to investigate how decentralized home-based networks of plug computers can support personal clouds according to sound architectural principles, mechanisms, and programming abstractions. To fulfill this vision we see three core scientific challenges, which we think must be overcome. The first challenge, decentralized churn-poor design, arises from the nature of plug federations, which show much lower levels of churn than traditional peer-to-peer environments. The second challenge, quasi-causal consistency, is caused by the simultaneous needs to produce a highly scalable environment (potentially numbering millions of users), that also offers collaborative editing capabilities of mutable data-structures (to offer rich social interactions). The third and final challenge, intuitive data structures for plug programming, arises from the need by programmers for intuitive and readily reusable data-structures to rapidly construct rich and robust decentralized personal cloud applications. This study is conducted in cooperation with other teams (GDD Team (University of Nantes), Inria team ASAP)
- **Labex COMINLAB contract (2014-2017): "Kharon-Security" - http://kharon.gforge.inria.fr**

  Google Play offers more than 800'000 applications (apps), and this number increases every day. Google play users have performed more than 25 billion app downloads. These applications vary from games to music, video, books, tools, etc. Unfortunately, each of these application is an attack vector on Android. The number of malicious applications (pieces of malware) discovered during the first six months of 2013 exceeds the number of pieces of malware discovered during the 2010 to 2012 period, more than 700 thousand malicious and risky applications were found in the wild. In this context, we propose the Kharon-Security project to stem the progression of Android pieces of malware. We propose to combine static and dynamic monitoring to compute a behavioral signature of Android malware. Behavioral signatures are helpful to understand how malware infect the devices and how they spread information in the Android operating system. Static analysis is essential to understand which particular event or callback triggers malware payload.

  In the project we have already developed GroddDroid a tool dedicated to automatic identification and execution of suspicious code. We have also built a dataset of Android malware, it this dataset, all malware are entirely manually reverse and documented. We have also developed an analysis platform. This platform is currently under private deployment.

- **Labex COMINLAB contract (2015-2018): "HardBlare-Security" - http://www.hardblare.cominlabs.ueb.eu/**

  The general context of the HardBlare project is to address Dynamic Information Flow Control that generally consists in attaching marks to denote the type of information that is saved or generated within the system. These marks are then propagated when the system evolves and information flow control is performed in order to guarantee a safe execution and storage within the system. Existing solutions imply a large overhead induced by the monitoring process. Some attempts rely on a hardware-software approach where DIFC operations are delegated to a coprocessor. Nevertheless, such approaches are based on modified processors. Beyond the fact hardware-assisted DIFC is hardly adopted, existing works do not take care of coprocessor security and multicore/multiprocessor embedded systems.

  We plan to implement DIFC mechanisms on boards including a non-modified ARM processor and a FPGA such as those based on the Xilinx Zynq family. The HardBlare project is a multidisciplinary project between CentraleSupélec IETR SCEE research team, CentraleSupélec Inria CIDRE research team and UBS Lab-STICC laboratory. Mounir Nasr Allah is doing his PhD in the context of this project. The main objective of this PhD is to study how hybrid analysis could improve hardware assisted DIFC using static analysis performed at compile-time. Another objective is to manage labels for persistent memory (i.e., files) using a modified OS kernel.

9.2. National Initiatives

9.2.1. ANR

- **ANR INFRA Project: SOCIOPLUG (2013-2017) - http://socioplug.univ-nantes.fr/index.php/SocioPlug_Project**

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

9.2.2. Inria Project Labs

- **CAPPRIS (2012-2016)**

  CAPPRIS stands for "Collaborative Action on the Protection of Privacy Rights in the Information Society". The main objective of CAPPRIS is to tackle the privacy challenges raised by the most recent developments and usages of information technologies such as profiling, data mining, social networking, location-based services or pervasive computing by developing solutions to enhance the protection of privacy in the Information Society. To solve this generic objective, the project focuses in particular on the following fundamental issues:

  - The design of appropriate metrics to assess and quantify privacy, primarily by extending and integrating the various possible definitions existing for the generic privacy properties such as anonymity, pseudonymity, unlinkability and unobservability, as well as notions coming from information theory or databases such as the recent but promising concept of differential privacy;

– The definition and the understanding of the fundamental principles underlying "privacy by design", with the hope of deriving practical guidelines to implement notions such as data minimization, proportionality, purpose specification, usage limitation, data sovereignty and accountability directly in the formal specifications of our information systems;
– The integration between the legal and social dimensions, intensely necessary since the developed privacy concepts, although they may rely on computational techniques, must be in adequacy with the applicable law (even in its heterogeneous and dynamic nature). In particular, privacy-preserving technologies cannot be considered efficient as long as they are not properly understood, accepted and trusted by the general public, an outcome which cannot be achieved by the means of a mathematical proof.

Three major application domains have been identified as interesting experimentation fields for this work: online social networks, location-based services and electronic health record systems. Each of these three domains brings specific privacy-related issues. The aim of the collaboration is to apply the techniques developed to the application domains in a way that promotes the notion of privacy by design, instead of simply considering them as a form of privacy add-ons on the top of already existing technologies. CAPPRIS is a joint project between Inria, LAAS-CNRS, Université de Rennes I, Supélec, Université de Namur, Eurecom, and Université de Versailles.

In addition of the scientific advances in the field of privacy, members of CAPPRIS are actively involved in the animation and federation of the French community on privacy, through the APVP workshop but also interdisciplinary colloquiums.

9.3. European Initiatives

9.3.1. FP7 & H2020 Projects

The **PANOPTESEC** project (http://www.panoptesec.eu) started on the 1st of November 2013 and ended in 2016. It deals with the automated and assisted security management of IT and SCADA system. The main objective of PANOPTESEC is to provide an integrated solution that will allow to efficiently monitor SCADA systems, detect intrusions and react to them. To that end, it encompasses many of the research topics that are addressed by the CIDRE team: alerts aggregation and correlation, policy-aware intrusion detection, architecture-aware intrusion detection, automated trust management, trust-based automated reaction and visualization.

The CIDRE team is involved in the project on all of these aspects. The partners are:

- REHA (BE),
- Nokia-Lucent Bell Labs France (FR),
- Epistematica (IT),
- the University of Rome (IT),
- the University of Hamburg (GE),
- the Institut Mines-Telecom (FR),
- ACEA (IT),
- CentraleSupélec (FR).

This year, our work focused on design and implementation but also on the integration phase. Most of our work focused on WP5 and WP6, that deal with the IDS event correlation system and the visualization system.

9.4. International Initiatives

9.4.1. Inria International Partners

9.4.1.1. Informal International Partners

Emmanuelle Anceaume is actively working with Leonardo Querzoni from the University La Sapienza, Italy, on data streams algorithms and engines. Their cooperation gave rise to two conference publications in 2016, one in Middelware [25] and the other one in Algotel [26].

Since several years, Michel Hurfin works with Professor Yun Wang (Southeast University, Nanjing, China). Their joint work focuses on convergence and synchronization problems in unreliable distributed systems prone to byzantine failures. In 2016, we investigate the iterative approximate byzantine consensus problem during a joint work with Chuanyou Li [22]. A visit of Professor Yun Wang in Rennes is planned next year.

9.5. International Research Visitors

9.5.1. Visits of International Scientists

Prof. Jean-Marc Robert from ETS (Ecole Supérieure de Technologie) of Montréal has made several short visits in the CIDRE research group in 2016.

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events Organisation

10.1.1.1. General Chair, Scientific Chair

Emmanuelle Anceaume served as the general chair of OPODIS 2015 (19th International Conference on Principles of Distributed Systems), December 2015, Rennes, France.

10.1.1.2. Member of the Organizing Committees

Christophe Bidan served as a member of the organization committee of C&ESAR 2016 (23rd Computers & Electronics Security Applications Rendez-vous), November 2016, Rennes, France.

Nicolas Prigent served as a member of the organization committee of SSTIC 2016 (Symposium sur la sécurité des technologies de l'information et des communications), June 2016, Rennes, France.

Frédéric Tronel served as a member of the organization committee of SSTIC 2016 (Symposium sur la sécurité des technologies de l'information et des communications), June 2016, Rennes, France.

10.1.2. Scientific Events Selection

10.1.2.1. Chair of Conference Program Committees

Nicolas Prigent serves as a program chair of VizSec 2016 (IEEE Symposium on Visualization for Cyber Security), October 2016, Baltimore, MD, USA.

10.1.2.2. Member of the Conference Program Committees

Emmanuelle Anceaume served as a member of the following program committees:

- Algotel 2016 (18èmes Rencontres Francophones sur les Aspects Algorithmiques des Télécommunications), May 2016, Bayonne, France.
- DSN 2016 (46th Annual IEEE/IFIP International Conference on Dependable Systems and Networks), June 2016, Toulouse, France.
- ATC 2016 (13th IEEE International Conference on Advanced and Trusted Computing), July 2016, Toulouse, France.
- PECCS 2016 (6th International Joint Conference on Pervasive and Embedded Computing and Communication Systems), July 2016, Lisbon, Portugal.
- TrustCom 2016 (5th IEEE International Conference on Trust, Security and Privacy in Computing and Communications, security track), August 2016, Tianjin, China.
- SRDS 2016 (35th International Symposium on Reliable Distributed Systems), September 2016, Budapest, Hungary.
- NCA 2016 (15th International Symposium on Network Computing and Applications), October 2016, Cambridge, MA, USA.

Christophe Bidan served as a member of the following program committees:

- CRiSIS 2016 (11th International Conference on Risks and Security of Internet and Systems), September 2016, Roscoff, France.
- C&ESAR 2016 (23rd Computers & Electronics Security Applications Rendez-vous), November 2016, Rennes, France.
- IWTCC 2016 (3rd International Workshop on Trust in Cloud Computing), December 2016, Shanghai, China.

Christopher Humphries served as a member of the program committee of VizSec 2016 (IEEE Symposium on Visualization for Cyber Security), October 2016, Baltimore, MD, USA.

Michel Hurfin served as a member of the following program committees:

- Ubisafe 2016 (8th IEEE International Symposium on UbiSafe Computing) August 2016, Tianjin, China.
- CARI 2016 (13rd African Conference on Research in Computer Science and Applied Mathematics), October 2016, Hammamet, Tunisia.

Frédéric Majorczyck served as a member of the program committee of VizSec 2016 (IEEE Symposium on Visualization for Cyber Security), October 2016, Baltimore, MD, USA.

Ludovic Mé served as a member of the following program committees:

- CARI 2016 (13rd African Conference on Research in Computer Science and Applied Mathematics), October 2016, Hammamet, Tunisia.
- AICCSA 2016 (13th ACS/IEEE International Conference on Computer Systems and Applications), November 2016, Agadir, Morocco.

Nicolas Prigent served as a member of the following program committees:

- SSTIC 2016 (Symposium sur la sécurité des technologies de l'information et des communications), June 2016, Rennes, France.
- GraMSec 2016 (3rd International workshop on Graphical Models for Security), June 2016, Lisbon, Portugal.

Eric Totel served as a member of the program committee of Ubisafe 2016 (8th IEEE International Symposium on UbiSafe Computing), August 2016, Tianjin, China.

Frédéric Tronel served as a member of the program committee of SSTIC 2016 (Symposium sur la sécurité des technologies de l'information et des communications) June 2016, Rennes, France.

10.1.2.3. Reviewer

- Laurent Georget - AICCSA 2016 (13th ACS/IEEE International Conference on Computer Systems and Applications).
- Gilles Guette - ICISSP 2016 (International Conference on Information System Security and Privacy) and AICCSA 2016 (13th ACS/IEEE International Conference on Computer Systems and Applications).
- Michel Hurfin - NCA 2016 (15th International Symposium on Network Computing and Applications).
- Ludovic Mé - NETYS 2016 (5th International Conference on NETworked sYStems).
- Valérie Viet Triem Tong - ICISSP 2016 (International Conference on Information Systems Security and Privacy).

10.1.3. Journal

10.1.3.1. Member of the Editorial Boards

Michel Hurfin belongs to the editorial board of the Springer open access journal of Internet Services and Applications.

10.1.3.2. Reviewer - Reviewing Activities

- Emmanuelle Anceaume - Elsevier JPDC (Journal of Parallel and Distributed Computing), Performance Evaluation, IEEE TDSC (Transactions on Dependable and Secure Computing), and IEEE TPDS (Transactions on Parallel and Distributed Systems).
- Michel Hurfin - Springer JISA (Journal of Internet Services and Applications) and Elsevier JPDC (Journal of Parallel and Distributed Computing).
- Ludovic Mé - journal "Revue Africaine de la Recherche en Informatique et Mathématique Appliquée".
- Guillaume Piolle - IJIS (International Journal on Information Security) and IEEE IS (Intelligent Systems).
- Valérie Viet Triem Tong - IEEE TPDS (Transactions on Parallel and Distributed Systems).

10.1.4. Invited Talks

Eric Totel has been invited to SEC2 (2nd workshop on Security in Clouds), July 2016, Lorient, France. He has given a talk about "Anomaly Based Intrusion Detection in Distributed Applications without global clock".

Valérie Viet Triem Tong has been invited to the Journées scientifiques Inria for a short presentation entitled Helping malware to execute themselves. She was also invited to give a talk during the Séminaire Aristote at the école Polytechnique (Palaiseau, France). Her presentation was concerning Android Malware analysis. Lastly, she has been invited at the 4th International Symposium on Information Systems Security (CISSI'2016 Morocco) for a invited talk about the need and the challenges in Education of formal computer programming.

10.1.5. Leadership within the Scientific Community

Ludovic Mé serves the Scientific Council of the LIRIMA (Laboratoire International de Recherche en Informatique et Mathématiques Appliquées).

Ludovic Mé chairs the steering Committee of the annual French conference RESSI (Rendez-Vous de la Recherche et de l'Enseignement de la Sécurité des Systèmes d'Information). He is a member of the Steering Committee of the annual international conference RAID (International Symposium on Research in Attacks, Intrusions and Defenses).

10.1.6. Scientific Expertise

Guillaume Piolle has been heard by the LIBE commission of the 82nd Internal Session of the European Youth Parliament, on "Data encryption, data protection and terrorism".

10.1.7. Research Administration

Emmanuelle Anceaume has participated in various juries (Post-doctoral grants, delegation Inria, PEDR Inria). As a member of the CE Inria, Emmanuelle Anceaume has participated to the hiring committee CR2/CR1 of Rennes and Sophia Antipolis.

Michel Hurfin is the local representative of the "mission jeunes chercheurs" in Rennes. He is a member of the "Commission personnel" and is in charge of the PhD student recruitment campaign of Inria Rennes Bretagne Atlantique. He is a member of the councils of the doctoral school Matisse. He is a member of the advisory board of the doctoral training center of EIT Digital in Rennes.

Ludovic Mé acts as Scientific Officer for the Rennes - Bretagne Atlantic Inria Research Center. As such, he is also a member of the Evaluation Commission and of the Internal Scientific Council of Inria.

Valérie Viet Triem Tong is member of the group working on the new Master proposal in computer science (for both CentraleSupelec and Inria)

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

Licence: Christophe Bidan, *Algorithms and Data Structures*, 36 hours of lecture including 7.5 hours of lectures, L3 - first year of the engineering degree, CentraleSupélec, France;

Licence: Christophe Bidan, *Software Engineering*, 12 hours, L3 - first year of the engineer degree, CentraleSupélec, France;

Licence: Christophe Bidan, *Supervision of student project*, 1 project, L3 - first year of the engineer degree, CentraleSupélec, France;

Master: Christophe Bidan is responsible for the module *Secured information systems*, M2 - third year of the engineer degree, CentraleSupélec, France;

Master: Christophe Bidan, *Applied cryptography*, 6 hours of lecture, M2 - master 2 degree, University of Rennes 1, France;

Master: Christophe Bidan, *Applied cryptography*, 15 hours including 6 hours of lecture, M2 - third year of the engineer degree, CentraleSupélec, France;

Master : Christophe Bidan, *Cryptographic Protocols*, 6 hours of lecture, mastère CS (Cyber Security), CentraleSupélec, France;

Master: Christophe Bidan, *Information systems*, 4.5 hours, M1 - second year of the engineer degree, CentraleSupélec, France;

Master: Christophe Bidan, *Supervision of student project*, 2 projects, M1 - second year of the engineer degree, CentraleSupélec, France;

Licence: Gilles Guette, *Network Initiation*, 57.5 hours, L3 - Licence, ISTIC/University of Rennes 1, France;

Licence: Gilles Guette, *Network Initiation*, 41.5 hours, L3 - first year of the engineer degree, ESIR, France;

Master: Gilles Guette, *Network Routing*, 32 hours, M1 - second year of the engineer degree, ESIR, France;

Master: Gilles Guette, *Mobile Network Routing*, 5 hours, M1 - second year of the engineer degree, ESIR, France;

Master: Gilles Guette, *Advanced Network Services*, 10 hours, M1 - second year of the engineer degree, ESIR, France;

Master: Gilles Guette, *Network Project*, 24 hours, M1 - second year of the engineer degree, ESIR, France;

Master: Gilles Guette, *Security*, 28 hours, M1 - second year of the engineer degree, ESIR, France;

Master: Gilles Guette, *Network and System Security*, 12 hours, M2 - third year of the engineer degree, ESIR, France;

Master: Gilles Guette, *Network Modeling*, 18 hours, M2 - third year of the engineer degree, ESIR, France;

Master: Gilles Guette, *Supervision of student project*, 1 project, M2 - third year of the engineer degree, ESIR, France;

Master: Gilles Guette, *Supervision of student internship*, M2 - Université Claude Bernard Lyon 1, France, Institut Francophone International, Hanoï, Viet-Nam;

Licence: Guillaume Hiet, *Algorithms and Data Structures*, 12.5 hours, L3 - first year of the engineer degree, CentraleSupélec, France;

Master: Guillaume Hiet, *Computer security and privacy for the engineer*, 8 hours, M1 - second year of the engineer degree, CentraleSupélec, France;

Master: Guillaume Hiet, *Buffer overflow vulnerabilities*, 16 hours, M2 - third year of the engineer degree, CentraleSupélec, France;

Master: Guillaume Hiet, *Buffer overflow vulnerabilities*, 16 hours, M2 - Mastère Spécialisé CS (Cyber Security), CentraleSupélec, France;

Master: Guillaume Hiet, *Pentest*, 19 hours, M2 - third year of the engineer degree, CentraleSupélec, France;

Master: Guillaume Hiet, *Pentest*, 3 hours, M2 - Mastère Spécialisé CS (Cyber Security), CentraleSupélec, France;

Master: Guillaume Hiet, *Introduction to Linux*, 3 hours, M2 - Mastère Spécialisé CS (Cyber Security), CentraleSupélec, France;

Master: Guillaume Hiet, *Java Security*, 4.5 hours, M2 - Mastère Spécialisé CS, CentraleSupélec, France;

Master: Guillaume Hiet, *Linux Security*, 18 hours, M2 - Mastère Spécialisé CS , CentraleSupélec, France;

Master: Guillaume Hiet, *Linux Security*, 7.5 hours, third year of the engineer degree, CentraleSupélec, France;

Master: Guillaume Hiet, *LDAP*, 7.5 hours, third year of the engineer degree, CentraleSupélec, France;

Master: Guillaume Hiet, *Intrusion Detection*, 15 hours, M2 - Mastère Spécialisé CS, CentraleSupélec, France;

Master: Guillaume Hiet, *Intrusion Detection*, 13.5 hours, M2 - third year of the engineer degree, M2 research degree of University of Rennes 1, CentraleSupélec, France;

Master: Guillaume Hiet, *Security Monitoring*, 3 hours, M2, cycle "Sécurité Numérique", INHESJ, France;

Master: Guillaume Hiet, *Computer Security*, 31.5 hours, M2, Mastère Spécialisé Architecte des Systèmes d'Information, CentraleSupélec, France;

Master: Guillaume Hiet, *Intrusion Detection*, 16 hours, M2, University of Rennes 1, France;

Master: Guillaume Hiet, *Intrusion Detection*, 10 hours, M2 - third year of the engineer degree, ESIR, France;

Master: Guillaume Hiet, *Intrusion Detection*, 9 hours, M2, Université of Limoges, France;

Master: Guillaume Hiet, *Firewall*, 6 hours, M2, University of Rennes 1, France;

Master: Guillaume Hiet, *Supervision of student project*, 4 projects, M1 - second year of the engineer degree, CentraleSupélec, France;

Master: Guillaume Hiet, *Supervision of student project*, 2 projects, M2 - third year of the engineer degree, CentraleSupélec, France;

Master: Guillaume Hiet, *Supervision of student project*, 2 projects, M2 - Mastère Spécialisé CS (Cyber Security), CentraleSupélec, France;

Licence : Ludovic Mé, *Software Engineering*, 18 hours, L3 - first year of the engineer degree, CentraleSupélec, France;

Licence : Ludovic Mé, *Software Engineering tutorials*, 6 hours, L3 - first year of the engineer degree, CentraleSupélec, France;

Licence : Ludovic Mé, *Software Engineering and Java development*, 18 hours, L3 - first year of the engineer degree, CentraleSupélec, France;

Master : Ludovic Mé, *Information systems tutorials*, 6 hours, M1 - second year of the engineer degree, CentraleSupélec, France;

Master : Ludovic Mé, *Operating systems tutorials*, 3 hours, M1 - second year of the engineer degree, CentraleSupélec, France;

Master : Ludovic Mé, *Supervision of student project*, 1 project, 38 hours, M1 - second year of the engineer degree, CentraleSupélec, France;

Licence : Guillaume Piolle, *Algorithms*, 16.5 hours, L3 - first year of the engineer degree, CentraleSupélec, France;

Licence : Guillaume Piolle, *Software engineering*, 18 hours, L3 - first year of the engineer degree, CentraleSupélec, France;

Master : Guillaume Piolle, *Modelling, Algorithms and Programming*, 30 hours, M1 - second year of the engineer degree, CentraleSupélec, France;

Master : Guillaume Piolle, *Computer security and privacy*, 9 hours, M1 - second year of the engineer degree, CentraleSupélec, France;

Master : Guillaume Piolle, *Computer networks*, 9 hours, M1 - second year of the engineer degree, CentraleSupélec, France;

Master : Guillaume Piolle, *Software project*, 8.5 hours, M1 - second year of the engineer degree, CentraleSupélec, France;

Master : Guillaume Piolle, *Security Policies*, 4.5 hours, M2 - third year of the engineer degree, CentraleSupélec, France;

Master : Guillaume Piolle, *Java programming*, 4.5 hours, M2 - third year of the engineer degree, CentraleSupélec, France;

Master : Guillaume Piolle, *Computer networks*, 9 hours, M2 - third year of the engineer degree, CentraleSupélec, France;

Master : Guillaume Piolle, *Software engineering*, 18 hours, M2 - third year of the engineer degree, CentraleSupélec, France;

Master : Guillaume Piolle, *Network Access Control*, 9 hours, M2 - third year of the engineer degree, CentraleSupélec, France;

Master : Guillaume Piolle, *Web development*, 18 hours, M2 - third year of the engineer degree, CentraleSupélec, France;

Master : Guillaume Piolle, *Privacy protection*, 4.5 hours, M2 - third year of the engineer degree, CentraleSupélec, France;

Master : Guillaume Piolle, *Computing project*, 40 hours, M2 - third year of the engineer degree, CentraleSupélec, France;

Licence : Eric Totel, *Models and programming languages*, 19.5 hours including 10.5 hours of lecture, L3 - first year of the engineer degree, CentraleSupélec, France;

Licence : Eric Totel, *Foundations of computer science, data structures and algorithms*, 6 hours, L3 - first year of the engineer degree, CentraleSupélec, France;

Licence : Eric Totel, *Software Modeling*, 15 hours, L3 - first year of the engineer degree, CentraleSupélec, France;

Master : Eric Totel, *Computer systems' architecture*, 60 hours, M1 - second year of the engineer degree, CentraleSupélec, France;

Master : Eric Totel, *C language*, 24 hours including 6 hours of lecture, M2 - master CS (Cyber Security), CentraleSupélec, France;

Master : Eric Totel, *C language and C++ language*, 12 hours including 6 hours of lecture, M2 - third year of the engineer degree, CentraleSupélec, France;

Master : Eric Totel, *Dependability* , 9 hours including 7.5 hours of lecture, M2 - third year of the engineer degree and master research, CentraleSupélec, France;

Master : Eric Totel, *Dependability*, 3 hours of lecture, M2 - third year of the engineer degree (ingénierie des systèmes automatisés), CentraleSupelec, France;

Master : Eric Totel, *Dependability*, 4.5 hours of lecture, M2 - post-graduate training (master Architecture des Réseaux de Communication), CentraleSupélec, France;

Master : Eric Totel, *Intrusion Detection*, 6 hours of lecture, M2 - M2 - master CS (Cyber Security), CentraleSupélec, France;

Master : Eric Totel, *Intrusion Detection*, 8 hours of lecture, M2 - master 2 degree, University of Rennes 1, France;

Master : Eric Totel, *Intrusion Detection*, 4 hours of lecture, M2 - master 2 degree, University of Rennes 1, France;

Master : Eric Totel, *Supervision of student project*, 4 projects, M1 - second year of the engineer degree, CentraleSupélec, France;

Master : Eric Totel, *Supervision of student project,* 1 project, M2 - third year of the engineer degree, CentraleSupélec, France;

Licence: Frédéric Tronel, *Software engineering*, 40 hours, L3 - first year of the engineer degree, CentraleSupélec, France;

Licence: Frédéric Tronel, *Operating Systems*, 12 hours, L3 - first year of the engineer degree, CentraleSupélec, France;

Master: Frédéric Tronel is responsible of the M2 degree in *CyberSecurity* (mastère spécialisé), organized jointly by CentraleSupélec and Télécom Bretagne, France;

Master: Frédéric Tronel, *Operating systems*, 21 hours hours, M2 - third year of the engineer degree, CentraleSupélec, France;

Master: Frédéric Tronel, *Compilers*, 18 hours, M2 - third year of the engineer degree, CentraleSupélec, France;

Master: Frédéric Tronel, *Automatic reasoning*, 4.5 hours, M2 - third year of the engineer degree, CentraleSupélec, France;

Master: Frédéric Tronel, *Assembly Language*, 6 hours, M2 - third year of the engineer degree, CentraleSupélec, France;

Master: Frédéric Tronel, *Buffer overflow vulnerabilities (theory and practice)*, 20.5 hours, M2 - third year of the engineer degree, CentraleSupélec, France;

Master: Frédéric Tronel, *Firewall*, 15 hours, M2 - third year of the engineer degree, CentraleSupélec, France;

Master: Frédéric Tronel, *Calculability in distributed systems*, 6 hours, M2, jointly with University of Rennes 1 and CentraleSupélec, France;

Master: Frédéric Tronel, *Computer network*, 8 hours, M2, jointly with University of Rennes 1 and CentraleSupélec, France;

Licence : Valérie Viet Triem Tong, *Algorithms and Data Structures*, 36 hours of lecture including 7 hours of lectures, L3 - first year of the engineering degree, CentraleSupélec, France;

Licence : Valérie Viet Triem Tong, *Supervision of student project*, 6 projects of 2nd year of the engineer degree, CentraleSupélec, France;

Master : Valérie Viet Triem Tong, *Games Theory*, 18 hours, M1 - second year of the engineering degree, CentraleSupélec, France;

Master : Valérie Viet Triem Tong, *Formal Methods*, 9 hours, M2 - third year of the engineering degree, CentraleSupélec, France;

Master : Valérie Viet Triem Tong, *Intrusion detection using information flow control*, 9 hours, M2 / third year of the engineering degree, CentraleSupélec, France;

Master : Valérie Viet Triem Tong, *Programming in Java*, 12 hours, M1 - international students (NplusI) second year of the engineering degree, CentraleSupélec, France;

Master : Valérie Viet Triem Tong, *Small elements of decidability*, 7.5 hours, M2 - third year of the engineering degree, CentraleSupélec, France;

Master : Valérie Viet Triem Tong, *Supervision of student project*, 1 project, mastere CS (Cyber Security), CentraleSupélec, France;

Master : Valérie Viet Triem Tong, *Supervision of student project*, 8 projects, M1 - second year of the engineer degree, CentraleSupélec, France;

Master : Valérie Viet Triem Tong, *Supervision of student project*, 2 projects mastere CS (Cyber Security), CentraleSupélec, France;

Master : Valérie Viet Triem Tong, *Supervision of student project*, 1 project year of the engineer degre, CentraleSupélec, France;

Doctorant : Valérie Viet Triem Tong, *Malware analysis by OS information flow tracking*, 2 hours, Inria, Summerschool - Cyber in Bretagne, France;

10.2.2. Supervision

PhD : Julien Lolive, *Entrelacement des mécanismes d'identification et de respect de la vie privée pour la protection des contenus externalisés*, May 2016, supervised by Caroline Fontaine (50% - Télécom-Bretagne) and Sébastien Gambs (50%);

PhD : Erwan Godefroy, *Définition et évaluation d'un mécanisme de génération de règles de corrélation liée à l'environnement*, September 2016, supervised by Michel Hurfin (33%), Eric Totel (33%), and Frédéric Majorczyk (34% - DGA MI);

PhD : Pierre Obame Meye, *Sûreté de fonctionnement dans le nuage de stockage*, December 2016, supervised by Emmanuelle Anceaume (33%), Frédéric Tronel (33%), and Philippe Raipin Parvedy (34% - Orange Labs);

PhD in progress: Deepak Subramanian, *Multi-level Information Flow Monitoring*, started in January 2013, supervised by Christophe Bidan (20%) and Guillaume Hiet (80%);

PhD in progress: Antoine Guellier, *Utilisation de la cryptographie homomorphe pour garantir le respect de la vie privée*, started in October 2013, supervised by Christophe Bidan (50%) and Nicolas Prigent (50%);

PhD in progress: Kun He, *Mise en œuvre de techniques de droit à l'oubli pour les contenus numériques*, started in October 2013, supervised by Christophe Bidan (50%) and Gae¨tan LeGuelvouit (50% - IRT B-Com);

PhD in progress: Mouna Hkimi, *Détection d'intrusion dans les systèmes distribués*, started in October 2013, supervised by Eric Totel (50%) and Michel Hurfin (50%);

PhD in progress: Solenn Brunet, *Privacy-preserving location-based services*, started in October 2014, supervised by Sébastien Gambs (50%) and Jacques Traoré (50% - Orange Labs Caen);

PhD in progress: Laurent Georget, *Validation Formelle d'un moniteur de flux d'information pour le noyau Linux*, started in October 2014, supervised by Mathieu Jaume (25% - MdC LIP6), Guillaume Piolle (25%), Frédéric Tronel (25%), and Valérie Viet Triem Tong (25%);

PhD in progress : Florian Grandhomme, *Protocole de routage externe type BGP dans un environnement réseaux tactiques adhoc mobiles : faisabilité, performances et qualité de service*, started in October 2014, supervised by Gilles Guette (50%), Adlen Ksentini (25% - Eurecom), and Thierry Plesse (25% - DGA MI);

PhD in progress: Thomas Letan, *Contribution à la sécurité des couches basses des systèmes d'information*, started in January 2015, supervised by Guillaume Hiet (50%), Pierre Chifflier (25% - ANSSI), and Ludovic Mé (25%);

PhD in progress: Oualid Koucham, *Détection d'intrusions pour les systèmes de contrôle industriels*, started in January 2015, supervised by Stéphane Mocanu (50% - Gipsa-lab), Guillaume Hiet (25%), and Jean-Marc Thiriet (25% - Gipsa-lab);

PhD in progress: Damien Crémilleux, *Visualisation d'évènements de sécurité pour la supervision*, started in October 2015, supervised by Christophe Bidan (30%), Nicolas Prigent (35%), and Frédéric Majorczyk (35% - DGA MI);

PhD in progress: Mourad Leslous, *Déclenchement automatique de codes jugés suspects dans les applications Android*, started in October 2015, supervised by Thomas Genet (20% - Celtique Inria project 20), Jean François Lalande (40% - INSA Centre Val de Loire), and Valérie Viet Triem Tong (40%);

PhD in progress: Mounir Nasr Allah, *Contrôle de flux d'information par utilisation conjointe d'analyse statique et d'analyse dynamique accélérée matériellement*, started in November 2015, supervised by Guillaume Hiet (75%) and Ludovic Mé (25%);

PhD in progress: Pernelle Mensah, *Adaptation de la Politique de Sécurité guidée par l'Evaluation du Risque dans les Infrastructures de Communication modernes*, started in January 2016, supervised by Eric Totel (25%), Guillaume Piolle (25%), Christine Morin (25% - Myriad Inria project), and Samuel Dubus (25% - Nokia);

PhD in progress: David Lanoë, *Détection d'intrusion dans les applications distribuées : l'approche comportementale comme alternative à la corrélation d'alertes*, started in october 2016, supervised by Michel Hurfin (50%) and Eric Totel (50%);

PhD in progress: Aurélien Trulla, *Caractérisation de malware Android par suivi de flux d'information et nouvelles techniques d'évasion*, started in October 2016, supervised by Jean Louis Lanet (25% - Tamis Inria project) and Valérie Viet Triem Tong (75%);

PhD in progress : Ronny Chevalier , "Enhanced computer platform security through an intrusion-detection approach", started in November 2016, supervised by Guillaume Hiet (50%), Boris Balacheff (25% - HP), and Ludovic Mé (25%);

PhD in progress: Laetitia Leichtnam, *Visualisation pour la caractérisation d'événements de sécurité*, started in october 2016, supervised by Eric Totel (40%), Nicolas Prigent (30%) and Ludovic Mé (30%);

PhD in progress : Charles Xosanavongsa, *Combining Attack Specification and Dynamic Learning from traces for correlation rule generation*, started in december 2016, supervised by Eric Totel (50%) and Ludovic Mé (50%);

PhD in progress : Yves Mocquard, *Population protocols*, started in september 2015, supervised by Bruno Sericola (Dyonisos Inria project) and Emmanuelle Anceaume.

10.2.3. Juries

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

- Antoine Rault, *Protection de la vie privée des utilisateurs dans un système de recommandations collaboratif distribué*, University of Rennes 1, 06/23/2016 (President of the Jury);
- Tarek Sayeh, '*Contrôle sélectif de l'accès à des données RDF*', University of Lyon 1, 09/08/2016 (President of the Jury);
- Ronan-Alexandre Cherrueau, *Composition de techniques de sécurité pour préserver la vie privée dans le contexte de l'informatique en nuage*, Ecole des Mines de Nantes, 11/18/2016 (Reviewer).

Eric Totel was a member of the PhD committee for the following PhD thesis:

- Siwar Kriaa, *Modélisation conjoint de la sûreté et de la sécurité pour l'évaluation des risques dans les systèmes cyber-physiques*, University of Paris-Saclay, Mars 2016 (President of the Jury);
- François Xavier Aguessy, *Evaluation Dynamique de Risque et Calcul de Réponses Basés sur des Modèles d'Attaques Bayésiens*, Telecom Sud-Paris and University Pierre & Marie Curie, September 2016. (Reviewer).

Guillaume Hiet was a member of the PhD committee for the PhD of Florent Marchand de Kerchov entitled *étendre des interpréteurs par détournement, ou comment étendre des interpréteurs sans en modifier le code*, prepared at Ecole des Mines de Nantes, 18 november 2016.

10.3. Popularization

Emmanuelle Anceaume was interviewed on the topic « Blockchain : comment le bitcoin révolutionne l'économie numérique ? » in a broadcast of "le labo des savoirs" (http://labodessavoirs.fr/emissions-du-labo/).

Guillaume Piolle has participated to the scientific popularization program *à la découverte de la recherche* aimed at secondary education pupils. His participation consisted in presentations about the objectives, methods and results of research activities in computer security and privacy (including, but not limited to our activities in CIDRE). It took place in high schools in Redon, Cesson-Sévigné and Dol-de-Bretagne.

Valérie Viet Triem Tong participates to (the stand and demo of) the 8h International Forum of CyberSecurity (FIC 2016).

Valérie Viet Triem Tong has published in Interstice a paper entitled "Lutter contre les codes malveillants" (https://interstices.info/jcms/p_91111/lutter-contre-les-codes-malveillants).

11. Bibliography

Publications of the year

Doctoral Dissertations and Habilitation Theses

[1] E. Godefroy.*Definition and assessment of a mechanism for the generation of environment-specific correlation rules*, CentraleSupélec, September 2016, https://hal.archives-ouvertes.fr/tel-01415703.

[2] J. Lolive.*Interleaving identification mechanisms and respect of privacy for the protection of outsourced content*, Télécom Bretagne, May 2016, https://hal.inria.fr/tel-01355495.

[3] P. O. Meye.*Dependability in Cloud Storage*, Université Rennes 1, December 2016, https://hal.archives-ouvertes.fr/tel-01413001.

Articles in International Peer-Reviewed Journal

[4] E. Anceaume, Y. Busnel, E. Schulte-Geers, B. Sericola.*Optimization Results for a Generalized Coupon Collector Problem*, in "Journal of Applied Probability", 2016, vol. 53, n^o 2, https://hal.inria.fr/hal-01397403.

[5] L. Caviglione, M. Gaggero, J.-F. Lalande, W. Mazurczyk, M. Urbanski.*Seeing the Unseen: Revealing Mobile Malware Hidden Communications via Energy Consumption and Artificial Intelligence*, in "IEEE Transactions on Information Forensics and Security", April 2016, vol. 11, n^o 4, p. 799-810 [*DOI :* 10.1109/TIFS.2015.2510825], https://hal.archives-ouvertes.fr/hal-01247495.

Invited Conferences

[6] E. Totel, M. Hkimi, M. Hurfin, M. Leslous, Y. Labiche.*Anomaly Based Intrusion Detection in Distributed Applications without global clock*, in "SEC2 2016 - Deuxième atelier sur la Sécurité dans les Clouds", Lorient, France, July 2016, https://hal.inria.fr/hal-01334608.

International Conferences with Proceedings

[7] M. ABDUL WAHAB, P. COTRET, M. NASR ALLAH, G. HIET, V. LAPOTRE, G. GOGNIAT.*Towards a hardware-assisted information flow tracking ecosystem for ARM processors*, in "26th International Conference on Field-Programmable Logic and Applications (FPL 2016)", Lausanne, Switzerland, August 2016, https://hal.archives-ouvertes.fr/hal-01337579.

[8] E. ANCEAUME, Y. BUSNEL, N. RIVETTI, B. SERICOLA.*Identifier des icebergs parmi des flux de données distribués*, in "ALGOTEL 2016 - 18èmes Rencontres Francophones sur les Aspects Algorithmiques des Télécommunications", Bayonne, France, ALGOTEL 2016 - 18èmes Rencontres Francophones sur les Aspects Algorithmiques des Télécommunications, May 2016, https://hal.archives-ouvertes.fr/hal-01303873.

[9] E. ANCEAUME, T. LAJOIE-MAZENC, R. LUDINARD, B. SERICOLA.*Safety Analysis of Bitcoin Improvement Proposals*, in "IEEE Symposium on Network Computing and Applications", Boston, United States, IEEE, October 2016, https://hal.archives-ouvertes.fr/hal-01397685.

[10] E. ANCEAUME, R. LUDINARD, B. SERICOLA.*L'empire romain ne doit pas être géré comme une petite île grecque*, in "ALGOTEL 2016 - 18èmes Rencontres Francophones sur les Aspects Algorithmiques des Télécommunications", Bayonne, France, May 2016, https://hal.archives-ouvertes.fr/hal-01305334.

[11] G. BROGI, V. VIET TRIEM TONG.*TerminAPTor: Highlighting Advanced Persistent Threats through Information Flow Tracking*, in "8th IFIP International Conference on New Technologies, Mobility and Security", Larnaca , Cyprus, November 2016, https://hal.inria.fr/hal-01417612.

[12] D. CRÉMILLEUX, C. BIDAN, F. MAJORCZYK, N. PRIGENT.*VEGAS: Visualizing, exploring and grouping alerts*, in "IEEE/IFIP International Workshop on Analytics for Network and Service Management", Istanbul, Turkey, April 2016, p. 1097 - 1100 [*DOI :* 10.1109/NOMS.2016.7502968], https://hal.archives-ouvertes.fr/hal-01416464.

[13] W. DE GROEF, D. SUBRAMANIAN, J. MARTIN, F. PIESSENS, D. LIEVEN.*Ensuring Endpoint Authenticity in WebRTC Peer-to-Peer Communication*, in "31st Annual ACM Symposium on Applied Computing (SAC 2016)", New York, United States, April 2016, https://hal.inria.fr/hal-01344572.

[14] F. GRANDHOMME, G. GUETTE, A. KSENTINI, T. PLESSE.*Comparing inter-domain routing protocol assessment tools for MANET*, in "2016 IEEE International Conference on Communications (ICC)", Kuala Lumpur, Malaysia, 2016 IEEE International Conference on Communications (ICC), May 2016, https://hal.inria.fr/hal-01355402.

[15] F. GRANDHOMME, G. GUETTE, A. KSENTINI, T. PLESSE.*ITMAN: An Inter Tactical Mobile Ad Hoc Network Routing Protocol*, in "MILCOM2016", Baltimore, United States, November 2016, https://hal.inria.fr/hal-01397710.

[16] K. HE, C. BIDAN, G. LE GUELVOUIT.*Experimentation of Privacy Protection for JPEG Contents on Image-Sharing Platforms*, in "9th International Conference on Security of Information and Networks (SIN 2016)", Rutgers University, New Jersey, United States, July 2016, https://hal.inria.fr/hal-01344469.

[17] K. HE, C. BIDAN, G. LE GUELVOUIT.*Privacy Protection for JPEG Content on Image-Sharing Platforms*, in "4th ACM Workshop on Information Hiding and Multimedia Security (IH&MMSEC 2016)", Vigo, Galicia, Spain, June 2016, https://hal.inria.fr/hal-01344472.

[18] K. HE, C. BIDAN, G. LE GUELVOUIT.*Robust and Secure Image Encryption Schemes During JPEG Compression Process*, in "2016 IS&T International Symposium on Electronic Imaging (EI 2016)", San Francisco, California, United States, February 2016, https://hal.inria.fr/hal-01344471.

[19] N. KISS, J.-F. LALANDE, M. LESLOUS, V. VIET TRIEM TONG.*Kharon dataset: Android malware under a microscope*, in "The Learning from Authoritative Security Experiment Results (LASER) workshop", San Jose, United States, USENIX Association, May 2016, p. 1-12, https://hal-centralesupelec.archives-ouvertes.fr/hal-01311917.

[20] O. KOUCHAM, S. MOCANU, G. HIET, J.-M. THIRIET, F. MAJORCZYK.*Detecting Process-Aware Attacks in Sequential Control Systems*, in "21st Nordic Conference on Secure IT Systems (NordSec 2016)", Oulu, Finland, November 2016, https://hal.inria.fr/hal-01361081.

[21] T. LETAN, P. CHIFFLIER, G. HIET, P. NÉRON, B. MORIN.*SpecCert: Specifying and Verifying Hardware-based Software Enforcement*, in "21st International Symposium on Formal Methods (FM 2016)", Limassol, Cyprus, 21st International Symposium on Formal Methods (FM 2016), Springer, November 2016, https://hal.inria.fr/hal-01361422.

[22] C. LI, M. HURFIN, Y. WANG, L. YU.*Towards a Restrained Use of Non-equivocation for Achieving Iterative Approximate Byzantine Consensus*, in "30th IEEE International Parallel and Distributed Processing Symposium (IPDPS)", Chicago, United States, May 2016, 10 [*DOI :* 10.1109/IPDPS.2016.62], https://hal.inria.fr/hal-01339477.

[23] Y. MOCQUARD, E. ANCEAUME, B. SERICOLA.*Optimal Proportion Computation with Population Protocols*, in "Symposium on Network Computing and Applications", Boston, United States, IEEE, October 2016, https://hal.archives-ouvertes.fr/hal-01354352.

[24] *Best Paper*
Y. MOCQUARD, B. SERICOLA, S. ROBERT, E. ANCEAUME.*Analysis of the Propagation Time of a Rumour in Large-scale Distributed Systems*, in "Symposium on Network Computing and Applications", Boston, United States, October 2016, This article has received the Best Student Paper Award, https://hal.archives-ouvertes.fr/hal-01354815.

[25] N. RIVETTI, E. ANCEAUME, Y. BUSNEL, L. QUERZONI, B. SERICOLA.*Online Scheduling for Shuffle Grouping in Distributed Stream Processing Systems Research Paper*, in "ACM/IFIP/USENIX Middleware 2016", Trento, Italy, ACM/IFIP/USENIX, December 2016 [*DOI :* 10.1145/2988336.2988347], https://hal.archives-ouvertes.fr/hal-01397658.

[26] N. RIVETTI, L. QUERZONI, E. ANCEAUME, Y. BUSNEL, B. SERICOLA.*Groupement de clés efficace pour un équilibrage de charge quasi-optimal dans les systèmes de traitement de flux*, in "ALGOTEL 2016 - 18èmes Rencontres Francophones sur les Aspects Algorithmiques des Télécommunications", Bayonne, France, ALGOTEL 2016 - 18èmes Rencontres Francophones sur les Aspects Algorithmiques des Télécommunications, May 2016, https://hal.archives-ouvertes.fr/hal-01303887.

[27] D. SUBRAMANIAN, G. HIET, C. BIDAN.*A self-correcting information flow control model for the web-browser*, in "The 9th International Symposium on Foundations & Practice of Security (FPS'2016)", Québec City, Canada, October 2016, https://hal.inria.fr/hal-01398192.

[28] *Best Paper*
D. SUBRAMANIAN, G. HIET, C. BIDAN.*Preventive Information Flow Control through a Mechanism of Split Addresses*, in "9th International Conference on Security of Information and Networks (SIN 2016)", Rutgers University, New Jersey, United States, July 2016, https://hal.inria.fr/hal-01344565.

[29] E. TOTEL, M. HKIMI, M. HURFIN, M. LESLOUS, Y. LABICHE.*Inferring a Distributed Application Behavior Model for Anomaly Based Intrusion Detection*, in "12th European Dependable Computing Conference", Gothenburg, Sweden, Proceedings of the 12th European Dependable Computing Conference, September 2016, https://hal.inria.fr/hal-01334596.

Conferences without Proceedings

[30] M. ABDUL WAHAB, P. COTRET, M. NASR ALLAH, G. HIET, V. LAPOTRE, G. COGNIAT.*A portable approach for SoC-based Dynamic Information Flow Tracking implementations*, in "11ème Colloque du GDR SoC/SiP", Nantes, France, June 2016, https://hal-centralesupelec.archives-ouvertes.fr/hal-01311045.

[31] E. ANCEAUME, R. LUDINARD, B. SERICOLA.*Relying on Consensus does not Make Bitcoin Safer*, in "Fast Abstract in the 46th Annual IEEE/IFIP International Conference on Dependable Systems and Networks", Toulouse, France, M. ROY, J. A. LOPEZ, A. CASIMIRO (editors), DSN2016-FAST-ABSTRACT, June 2016, https://hal.archives-ouvertes.fr/hal-01316541.

[32] M. ASSAF, J. SIGNOLES, E. TOTEL, F. TRONEL.*The Cardinal Abstraction for Quantitative Information Flow*, in "Workshop on Foundations of Computer Security 2016 (FCS 2016)", Lisbon, Portugal, June 2016, https://hal.inria.fr/hal-01334604.

[33] F. GRANDHOMME, G. GUETTE, A. KSENTINI, T. PLESSE.*Comparaison d'outils d'évaluation de performance des protocoles de routage inter-MANET*, in "13ème Conference Francophone sur les Nouvelles Technologies de la Répartition (NOTERE 2016)", Paris, France, July 2016, https://hal.inria.fr/hal-01355407.

[34] P. MENSAH.*Dynamic Topology Extraction in Cloud Infrastructures*, in "Second workshop on Security in Clouds (SEC2)", Lorient, France, July 2016, https://hal.inria.fr/hal-01399251.

Research Reports

[35] E. ANCEAUME, F. CASTELLA, A. MOSTEFAOUI, B. SERICOLA.*Performance Evaluation of a Distributed Synchronization Protocol*, Inria ; Irisa ; Lina ; Irmar, March 2016, https://hal.inria.fr/hal-01283064.

[36] G. BONNORON, D. CRÉMILLEUX, S. T. BULUSU, X. ZHU, G. VALADON.*Survey and analysis of DNS infrastructures*, CNRS, 2016, https://hal.archives-ouvertes.fr/hal-01407640.

[37] T. LETAN, P. CHIFFLIER, G. HIET, P. NÉRON, B. MORIN.*SpecCert: Specifying and Verifying Hardware-based Security Enforcement*, CentraleSupélec ; Agence Nationale de Sécurité des Systèmes d'Information, 2016, 20, https://hal.inria.fr/hal-01356690.

[38] N. RIVETTI, E. ANCEAUME, Y. BUSNEL, L. QUERZONI, B. SERICOLA.*Load-Aware Shedding in Stream Processing Systems*, LINA-University of Nantes ; Sapienza Università di Roma (Italie) ; Irisa ; Inria Rennes, May 2016, https://hal.inria.fr/hal-01311970.

Scientific Popularization

[39] V. VIET TRIEM TONG, J.-F. LALANDE, M. LESLOUS.*Challenges in Android Malware Analysis*, in "ERCIM News", July 2016, n^{o} 106, p. 42-43, https://hal-centralesupelec.archives-ouvertes.fr/hal-01355122.

[40] V. VIET TRIEM TONG.*Lutter contre les codes malveillants*, in "Interstices", December 2016, https://hal.inria.fr/hal-01427326.

Other Publications

[41] M. HKIMI.*Apprentissage d'un modèle comportemental d'une application distribuée pour la détection d'intrusion*, May 2016, Rendez-Vous de la Recherche et de l'enseignement de la Sécurité des Systèmes d'Information (RESSI 2016), Poster, https://hal.inria.fr/hal-01334612.

[42] M. LESLOUS, J.-F. LALANDE, V. VIET TRIEM TONG.*Using Implicit Calls to Improve Malware Dynamic Execution*, May 2016, 37th IEEE Symposium on Security and Privacy, Poster, https://hal.archives-ouvertes.fr/hal-01304326.

[43] M. NASR ALLAH, G. HIET, M. ABDUL WAHAB, P. COTRET, G. GOGNIAT, V. LAPOTRE.*HardBlare: a Hardware-Assisted Approach for Dynamic Information Flow Tracking*, April 2016, Séminaire des doctorantes et doctorants en informatique de la Société Informatique de France, Poster, https://hal-centralesupelec.archives-ouvertes.fr/hal-01311032.

Project-Team DIONYSOS

Dependability Interoperability and perfOrmance aNalYsiS Of networkS

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

IN PARTNERSHIP WITH:
Université Rennes 1

RESEARCH CENTER
Rennes - Bretagne-Atlantique

THEME
Networks and Telecommunications

Table of contents

Project-Team DIONYSOS

Creation of the Project-Team: 2009 January 01

Keywords:

Computer Science and Digital Science:

1.1.6. - Cloud
1.1.7. - Peer to peer
1.1.13. - Virtualization
1.2.3. - Routing
1.2.4. - QoS, performance evaluation
1.2.5. - Internet of things
1.3. - Distributed Systems
3.4.1. - Supervised learning
3.4.2. - Unsupervised learning
3.4.3. - Reinforcement learning
3.4.6. - Neural networks
3.4.8. - Deep learning
6.1.1. - Continuous Modeling (PDE, ODE)
6.2.2. - Numerical probability
6.2.3. - Probabilistic methods
6.2.6. - Optimization
7.2. - Discrete mathematics, combinatorics
7.14. - Game Theory

Other Research Topics and Application Domains:

1.3.1. - Understanding and simulation of the brain and the nervous system
2.2. - Physiology and diseases
6.2.1. - Wired technologies
6.2.2. - Radio technology
6.2.4. - Optic technology
6.3.2. - Network protocols
6.4. - Internet of things

1. Members

Research Scientists

Gerardo Rubino [Team leader, Inria, Senior Researcher, HDR]
Bruno Sericola [Inria, Senior Researcher, HDR]
Bruno Tuffin [Inria, Senior Researcher, HDR]

Faculty Members

Pierre L'Ecuyer [Inria, Professor]
Yassine Hadjadj-Aoul [Univ. Rennes I, Associate Professor]
Sofiene Jelassi [Univ. Rennes I, Associate Professor, since September 2016]
Raymond Marie [Univ. Rennes I, Professor, HDR]

César Viho [Univ. Rennes I, Professor, HDR]
Adlen Ksentini [Univ. Rennes I, Associate Professor, until March 2016, HDR]
Yann Busnel [External member, Associate Professor at ENSAI, Rennes, HDR]

Technical Staff

Pantelis Frangoudis [Univ. Rennes I]
Federico Sismondi [Univ. Rennes I]
Thiago Wanderley Matos de Abreu [Univ. Rennes I, until Aug 2016]

PhD Students

Imad Alawe [TDF, granted by CIFRE]
Hamza Ben Ammar [Univ. Rennes I]
Corentin Hardy [Technicolor, granted by CIFRE]
Nicolas Jara Carvallo [Scolarship from Chilian gouv.]
Farouk Messaoudi [Inst. de Recherche Technologique B-COM]
Yves Mocquard [Univ. Rennes I]
Quang Pham Tran Anh [Univ. Rennes I]
Ajit Rai [ALSTOM, granted by CIFRE]
Ali Hodroj [Univ. Rennes I and Saint Joseph Univ. (Lebanon)]
Louiza Yala [Univ. Rennes I]
Osama Arouk [Univ. Rennes I, until March 2016]
Jorge Graneri [Univ. of the Republic, Uruguay]
Laura Aspirot [Univ. of the Republic, Uruguay]
Btissam Er-Rahmadi [Orange Lab., granted by CIFRE, until Jun 2016]
Yue Li [Orange Lab., granted by CIFRE]
Jean-Michel Sanner [Orange Lab., granted by CET]

Visiting Scientists

Nawel Attia [Faculté des Sciences de Tunis ELMANAR, until Oct 2016]
Marta Barria [Univ. of Valparaíso, Chile, July 2016]
Meriem Bouzouita [Supcom, Tunisia]
Hector Cancela [Univ. of the Republic, Uruguay, 2 weeks in Dec. 2016]
Yassine Douga [Univ. of Oran, Algeria, until Apr 2016]
Claudio Risso [Univ. of the Republic, Uruguay, 2 weeks in Dec. 2016]
Franco Robledo [Univ. of the Republic, Uruguay, 2 weeks in Dec. 2016]
Reinaldo Vallejos [UTFSM, Chile, July 2016]

Administrative Assistant

Fabienne Cuyollaa [Inria]

Others

Mulugeta Tamiru [Univ. Rennes I, from May 2016 until Aug 2016]
Tompoariniaina Andriamilanto [Univ. Rennes I, from May 2016 until Aug 2016]

2. Overall Objectives

2.1. Overall objectives

The main objectives of the project are the identification, the conception and the selection of the most appropriate network architectures for a communication service, as well as the development of computing and mathematical tools for the fulfillment of these tasks. These objectives lead to two types of complementary research fields: the systems' qualitative aspects (e.g. protocol testing and design) and the quantitative aspects which are essential to the correct dimensioning of these architectures and the associated services (performance, dependability, Quality of Service (QoS), Quality of Experience (QoE) and performability); our activities lie essentially in the latter.

The Dionysos group works on different problems related to the design and the analysis of communication services. Such services require functionality specifications, decisions about where and how they must be deployed in a system, and the dimensioning of their different components. The interests of the project concern not only particular classes of systems but also methodological aspects.

Concerning the communication systems themselves, we focus on IP networks, at different levels. Concerning the types of networks considered, we mainly work in the wireless area, in particular on sensor networks, on Content Delivery Networks for our work around measuring the perceived quality, the main component of QoE, and on some aspects of optical networks. We also work on the assessment of interoperability between specific network components, which is essential to ensure that they interact correctly before they get deployed in a real environment. Our team contributes in providing solutions (methods, algorithms and tools) which help in obtaining efficient interoperability test suites for new generation networks. From the application point of view, we also have activities in network economics methodologies, a critical multi-disciplinary area for telecommunications providers, with many defying open problems for the near future.

For most of previous mentioned problems, our work concern their quantitative aspects. The quantitative aspects we are interested in are QoE, performance, dependability, performability, QoS, vulnerability, etc. We develop techniques for the evaluation of these different aspects of the considered systems through *models* and through *measurement techniques*. In particular, we develop techniques to measure in an automatic way the quality of a video or audio communication *as perceived by the final user*. The methods we work with range from discrete event simulation and Monte Carlo procedures to analytical techniques, and include numerical algorithms as well. Our main mathematical tools are stochastic processes in general and queuing models and Markov chains in particular, optimization techniques, graph theory, combinatorics, etc.

3. Research Program

3.1. Introduction

The scientific foundations of our work are those of network design and network analysis. Specifically, this concerns the principles of packet switching and in particular of IP networks (protocol design, protocol testing, routing, scheduling techniques), and the mathematical and algorithmic aspects of the associated problems, on which our methods and tools are based.

These foundations are described in the following paragraphs. We begin by a subsection dedicated to Quality of Service (QoS) and Quality of Experience (QoE), since they can be seen as unifying concepts in our activities. Then we briefly describe the specific sub-area of model evaluation and about the particular multidisciplinary domain of network economics.

3.2. Quality of Service and Quality of Experience

Since it is difficult to develop as many communication solutions as possible applications, the scientific and technological communities aim towards providing general *services* allowing to give to each application or user a set of properties nowadays called “Quality of Service” (QoS), a terminology lacking a precise definition. This QoS concept takes different forms according to the type of communication service and the aspects which matter for a given application: for performance it comes through specific metrics (delays, jitter, throughput, etc.), for dependability it also comes through appropriate metrics: reliability, availability, or vulnerability, in the case for instance of WAN (Wide Area Network) topologies, etc.

QoS is at the heart of our research activities: We look for methods to obtain specific “levels” of QoS and for techniques to evaluate the associated metrics. Our ultimate goal is to provide tools (mathematical tools and/or algorithms, under appropriate software “containers” or not) allowing users and/or applications to attain specific levels of QoS, or to improve the provided QoS, if we think of a particular system, with an optimal use of the resources available. Obtaining a good QoS level is a very general objective. It leads to many different areas, depending on the systems, applications and specific goals being considered. Our team works on several of these areas. We also investigate the impact of network QoS on multimedia payloads to reduce the impact of congestion.

Some important aspects of the behavior of modern communication systems have subjective components: the quality of a video stream or an audio signal, *as perceived by the user*, is related to some of the previous mentioned parameters (packet loss, delays, ...) but in an extremely complex way. We are interested in analyzing these types of flows from this user-oriented point of view. We focus on the *user perceived quality*, in short, PQ, the main component of what is nowadays called Quality of Experience (in short, QoE), to underline the fact that, in this case, we want to center the analysis on the user. In this context, we have a global project called PSQA, which stands for Pseudo-Subjective Quality Assessment, and which refers to a technology we have developed allowing to automatically measure this PQ.

Another special case to which we devote research efforts in the team is the analysis of qualitative properties related to interoperability assessment. This refers to the act of determining if end-to-end functionality between at least two communicating systems is as required by the base standards for those systems. Conformance is the act of determining to what extent a single component conforms to the individual requirements of the standard it is based on. Our purpose is to provide such a formal framework (methods, algorithms and tools) for interoperability assessment, in order to help in obtaining efficient interoperability test suites for new generation networks, mainly around IPv6-related protocols. The interoperability test suites generation is based on specifications (standards and/or RFCs) of network components and protocols to be tested.

3.3. Stochastic modeling

The scientific foundations of our modeling activities are composed of stochastic processes theory and, in particular, Markov processes, queuing theory, stochastic graphs theory, etc. The objectives are either to develop numerical solutions, or analytical ones, or possibly discrete event simulation or Monte Carlo (and Quasi-Monte Carlo) techniques. We are always interested in model evaluation techniques for dependability and performability analysis, both in static (network reliability) and dynamic contexts (depending on the fact that time plays an explicit role in the analysis or not). We look at systems from the classical so-called *call level*, leading to standard models (for instance, queues or networks of queues) and also at the *burst level*, leading to *fluid models*.

In recent years, our work on the design of the topologies of WANs led us to explore optimization techniques, in particular in the case of very large optimization problems, usually formulated in terms of graphs. The associated methods we are interested in are composed of simulated annealing, genetic algorithms, TABU search, etc. For the time being, we have obtained our best results with GRASP techniques.

Network pricing is a good example of a multi-disciplinary research activity half-way between applied mathematics, economy and networking, centered on stochastic modeling issues. Indeed, the Internet is facing a tremendous increase of its traffic volume. As a consequence, real users complain that large data transfers take too long, without any possibility to improve this by themselves (by paying more, for instance). A possible solution to cope with congestion is to increase the link capacities; however, many authors consider that this is not a viable solution as the network must respond to an increasing demand (and experience has shown that demand of bandwidth has always been ahead of supply), especially now that the Internet is becoming a commercial network. Furthermore, incentives for a fair utilization between customers are not included in the current Internet. For these reasons, it has been suggested that the current flat-rate fees, where customers pay a subscription and obtain an unlimited usage, should be replaced by usage-based fees. Besides, the future Internet will carry heterogeneous flows such as video, voice, email, web, file transfers and remote login among others. Each of these applications requires a different level of QoS: for example, video needs very small delays and packet losses, voice requires small delays but can afford some packet losses, email can afford delay (within a given bound) while file transfer needs a good average throughput and remote login requires small round-trip times. Some pricing incentives should exist so that each user does not always choose the best QoS for her application and so that the final result is a fair utilization of the bandwidth. On the other hand, we need to be aware of the trade-off between engineering efficiency and economic efficiency; for example, traffic measurements can help in improving the management of the network but is a costly option. These are some of the various aspects often present in the pricing problems we address in our work. More recently, we have switched to the more general field of network economics, dealing with the economic behavior of users, service providers and content providers, as well as their relations.

4. Application Domains

4.1. Networking

Our global research effort concerns networking problems, both from the analysis point of view, and around network design issues. Specifically, this means the IP technology in general, with focus on specific types of networks seen at different levels: wireless systems, optical infrastructures, peer-to-peer architectures, Software Defined Networks, Content Delivery Networks, Content-Centric Networks, clouds.

A specific aspect of network applications and/or services based on video or voice content, is our PSQA technology, able to measure the Perceptual Quality automatically and in real time. PSQA provides a MOS value as close as it makes sense to the value obtained from subjective testing sessions. The technology has been tested in many environments, including one way communications as, for instance, in video streaming, and bi-directional communications as in IP telephony, UDP- or TCP-based systems, etc. It has already served in many collaborative projects as the measuring tool used.

4.2. Stochastic modeling

Many of the techniques developed at Dionysos are related to the analysis of complex systems in general, not only in telecommunications. For instance, our Monte Carlo methods for analyzing rare events have been used by different industrial partners, some of them in networking but recently also by companies building transportation systems. We develop methods in different areas: numerical analysis of stochastic models, bound computations in the same area, Discrete Event Simulation, or, as just mentioned, rare event analysis.

5. Highlights of the Year

5.1. Highlights of the Year

Pierre L'Ecuyer received the 2016 ACM SIGSIM Distinguished Contributions Award.
BEST PAPERS AWARDS :
[75] **Symposium on Network Computing and Applications**. Y. MOCQUARD, B. SERICOLA, S. ROBERT, E. ANCEAUME.
[49] **International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems (MSWIM 2016)**. M. BOUZOUITA, Y. HADJADJ-AOUL, N. ZANGAR, G. RUBINO, S. TABBANE.

6. New Software and Platforms

6.1. The Passive Test Tool (ttproto) and CoAP Testing Tool

`ttproto` is an experimental tool for implementing testing tools, for conformance and interoperability testing. It was first implemented to explore interesting features and concepts for the TTCN-3 standard. It was also used to implement a passive interoperability test suite we provided for the CoAP (Constrained Application Protocol) interoperability event held in Paris in March 2012. It is currently used for the purpose of developing testing tools (for interoperability and conformance testing) for the F-interop european project (see http://www.f-interop.eu/). This tool is implemented in `python3` and its design was influenced mainly by TTCN-3 (abstract model, templates, snapshots, behavior trees, communication ports, logging) and by Scapy (syntax, flexibility, customizability) Its purpose is to facilitate rapid prototyping and experimentation (rather than production use). We chose to maximize its modularity and readability rather than performances and real-time considerations.

- Participants: César Viho, Federico Sismondi
- Contact: César Viho, Federico Sismondi
- http://www.irisa.fr/tipi

6.2. T3DevKit and IPv6 testing tools

We have built a toolkit for easing executing tests written in the standardized TTCN-3 test specification language. This toolkit is made of a C++ library together with a highly customizable CoDec generator that allows fast development of external components (that are required to execute a test suite) such as CoDec (for message Coding/Decoding), System and Platform Adapters. It also provides a framework for representing and manipulating TTCN-3 events so as to ease the production of test reports. The toolkit addresses issues that are not yet covered by ETSI standards while being fully compatible with the existing standard interfaces: TRI (Test Runtime Interfaces) and TCI (Test Control Interfaces), it has been tested with four TTCN-3 environments (IBM, Elvior, Danet and Go4IT) and on three different platforms (Linux, Windows and Cygwin).

T3DevKit is a free open source toolkit to ease the development of test suites in the TTCN-3 environment. It provides a CoDec generator (t3cdgen) that automates the development process of the CoDec needed for coding TTCN-3 values into physically transmittable messages and decoding incoming messages. A library (t3devlib) provides an object oriented framework to manipulate TTCN-3 entities (values, ports, timers, external functions). T3DevKit offers an implementation of the TRI and TCI standard interfaces a default implementations for the system adapter (SA), platform adapter (PA), test management (TM), test logging (TL) and component handling (CH) modules and a default codec. Built-in scripts for the generation of executable test suite, which are tool-independent facilitate the distribution of test suite sources.

IPv6 test suites have been developed using the TTCN-3 environment. The full Abstract Test Suites are written in TTCN-3 and the source files for building the codecs and adapters with the help of T3DevKit.

- Participants: César Viho, Federico Sismondi
- Contact: Federico Sismondi

6.3. Performance and dependability evaluation

Participants: Gerardo Rubino, Bruno Sericola, Bruno Tuffin.

We develop software tools for the evaluation of two classes of models: Markov models and reliability networks. The main objective is to quantify dependability aspects of the behaviors of the modeled systems, but other aspects of the systems can be handled (performance, performability, vulnerability). The tools are specialized libraries implementing numerical, Monte Carlo and Quasi-Monte Carlo algorithms.

One of these libraries has been developed for the Celar (DGA), and its goal is the evaluation of dependability and vulnerability metrics of wide area communication networks (WANs). The algorithms in this library can also evaluate the sensitivities of the implemented dependability measures with respect to the parameters characterizing the behavior of the components of the networks (nodes, lines).

We are also developing tools with the objective of building Markovian models and to compute bounds of asymptotic metrics such as the asymptotic availability of standard metrics of models in equilibrium, loss probabilities, blocking probabilities, mean backlogs, etc. A set of functions designed for dependability analysis is being built under the internal name `DependLib`.

We contribute to the development of SPNP (*Stochastic Petri Net Package*). SPNP is used by more than 200 companies and universities. The main designer is Duke University. Our contributions are on Monte Carlo methods. We plan to increase our participation in the development of this tool.

Pierre L'Ecuyer is also developing in Montreal a library, *Stochastic Simulation in Java* (SSJ), providing facilities for generating uniform and nonuniform random variates, computing different measures related to probability distributions, performing goodness-of-fit tests, applying quasi-Monte Carlo methods, collecting (elementary) statistics, and programming discrete-event simulations with both events and processes.

7. New Results

7.1. Performance Evaluation of Call Centers

Participant: Pierre L'Ecuyer.

We develop research activities around the analysis and design of call centers, from a performance perspective. The effective management of call centers is a challenging task mainly because managers are consistently facing considerable uncertainty.

One aspect studied in [23] is the development of stochastic models for the daily arrival rate in a call center. Models in which the busyness factors are independent across periods, or in which a common busyness factor applies to all periods, have been studied previously. But they are not sufficiently realistic. We examine alternative models for which the busyness factors have some form of dependence across periods.

We also carry out in [14] large-scale data-based investigation of service times in a call center with many heterogeneous agents and multiple call types to investigate the validity of traditionally used standard Erlang queueing models, based on independent and identically distributed exponential random variables. Our study provides empirical support to the theoretical research that goes beyond standard modelling assumptions in service systems.

In [56], we consider a stochastic staffing problem with uncertain arrival rates. The objective is to minimize the total cost of agents under some chance constraints, defined over the randomness of the service level in a given time period. We present a method that combines simulation, mixed integer programming, and cut generation to solve this problem. In [84], we consider a particular staffing problem with probabilistic constraints in an emergency call center. We propose an algorithm to solve the problem, and validate it with a simulation model based on real data from the 911 emergency call center of Montreal, Canada.

We are also interested in predicting the waiting time of customers upon their arrival in some service system such as a call center or emergency service. In [86], we propose two new predictors that are very simple to implement and can be used in multiskill settings. They are based on the waiting times of previous customers of the same class. In our simulation experiments, these new predictors are very competitive with the optimal ones for a simple queue, and for multiskill centers they perform better than other predictors of comparable simplicity.

7.2. Analytic models

Participants: Gerardo Rubino, Bruno Sericola.

Sojourn times in Markovian models. In [98], we discuss different issues related to the time a Markov chain spends in a part of its state space. This is relevant in many application areas including those interesting Dionysos, namely, in the performance and dependability analysis of complex systems. For instance, in dependability, the reliability of a system subject to failures and repairs of its components, is, in terms of a discrete-space model of it, the probability that it remains in the subset of operational or up states during the whole time interval $[0, t]$. In performance, the occupancy factor of some server is the probability that, in steady state, the model belongs to the subset of states where the server is busy. This book chapter reviews some past work done by the authors on this topic (see our book [111] for a synthesis of these works), and add some new insights on the properties of these sojourn times.

Queuing systems in equilibrium. In the late 70s, Leonard Kleinrock proposed a metric able to capture the tradeoff between the work done by a system and its cost, or, in terms of queueing systems, between throughput and mean response time. The new metric was called *power* and among its properties, it satisfies a nice one informally called "keep the pipe full", specifying that the operation point of many queues that maximizes their power also leads to a mean backlog equal to exactly one customer. Last year [110] we explored what happens with this metric when we consider Jackson queuing networks. After showing that the same property holds for them, we showed that the power metric has some drawbacks, mainly when considering multiserver queues and networks of queues. We then proposed a new metric that we called *effectiveness*, identical to power when there is a single queue with a single server, but different otherwise, that avoids these drawbacks. We analyze it and, in particular, we showed that the same "keep the pipe full" holds for it. In the keynote [34] we presented these ideas together with some new results (for example, the analysis of G-queues from this point of view).

For other analytical-oriented work, see [72] for new applications of queueing theory used at the Markovian level, and [72] for applications of stochastic analysis to general problems where performance and dependability are simultaneously taken into account in the same model.

7.3. Performance Evaluation of Distributed Systems

Participants: Bruno Sericola, Yann Busnel, Yves Mocquard.

Detection of distributed deny of service attacks. A Deny 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 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, inform a monitoring server that aggregates all the monitored information to accurately detect icebergs [41]. 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 (modeled as 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 [10].

Stream Processing Systems. Stream processing systems are today gaining momentum as tools to perform analytics on continuous data streams. Their ability to produce analysis results with sub-second latencies, coupled with their scalability, makes them the preferred choice for many big data companies.

A stream processing application is commonly modeled as a direct acyclic graph where data operators, represented by nodes, are interconnected by streams of tuples containing data to be analyzed, the directed edges (the arcs). Scalability is usually attained at the deployment phase where each data operator can be parallelized using multiple instances, each of which will handle a subset of the tuples conveyed by the operators' ingoing stream. Balancing the load among the instances of a parallel operator is important as it yields to better resource utilization and thus larger throughputs and reduced tuple processing latencies. We have proposed a new key grouping technique targeted toward applications working on input streams characterized by a skewed value distribution [80]. Our solution is based on the observation that when the values used to perform the grouping have skewed frequencies, the few most frequent values (the *heavy hitters*) drive the load distribution, while the remaining largest fraction of the values (the *sparse items*) appear so rarely in the stream that the relative impact of each of them on the global load balance is negligible. We have shown, through a theoretical analysis, that our solution provides on average near-optimal mappings using sub-linear spaces in the number of tuples read from the input stream in the learning phase and the support (value domain) of the tuples. In particular this analysis presents new results regarding the expected error made on the estimation of the frequency of heavy hitters.

Load shedding is a technique employed by stream processing systems to handle unpredictable spikes in the input load whenever available computing resources are not adequately provisioned. A load shedder drops tuples to keep the input load below a critical threshold and thus avoid unbounded queuing and system trashing.

In [102] and [79] we propose Load-Aware Shedding (LAS), a novel load shedding solution that, unlike previous works, does not rely neither on a pre-defined cost model nor on any assumption on the tuple execution duration. Leveraging sketches, LAS efficiently builds and maintains at runtime a cost model to estimate the execution duration of each tuple with small error bounds. This estimation enables a proactive load shedding of the input stream at any operator that aims at limiting queuing latencies while dropping as few tuples as possible. We provide a theoretical analysis proving that LAS is an (ε, δ)-approximation of the optimal online load shedder. Furthermore, through an extensive practical evaluation based on simulations and a prototype, we evaluate its impact on stream processing applications, which validate the robustness and accuracy of LAS.

Shuffle grouping is a technique used by stream processing frameworks to share input load among parallel instances of stateless operators. With shuffle grouping each tuple of a stream can be assigned to any available operator instance, independently from any previous assignment. A common approach to implement shuffle grouping is to adopt a Round-Robin policy, a simple solution that fares well as long as the tuple execution time is almost the same for all the tu-ples. However, such an assumption rarely holds in real cases where execution time strongly depends on tuple content. As a consequence, parallel stateless operators within stream processing applications may experience unpredictable unbalance that, in the end, causes undesirable increase in tuple completion times. In [77] we propose Online Shuffle Grouping (OSG), a novel approach to shuffle grouping aimed at reducing the overall tuple completion time. OSG estimates the execution time of each tuple, enabling a proactive and online scheduling of input load to the target operator instances. Sketches are used to efficiently store the otherwise large amount of information required to schedule incoming load. We provide a probabilistic analysis and illustrate, through both simulations and a running prototype, its impact on stream processing applications.

Estimating the frequency of any piece of information in large-scale distributed data streams became of utmost importance in the last decade (*e.g.*, in the context of network monitoring, big data, *etc.*). If some elegant solutions have been proposed recently, their approximation is computed from the inception of the stream. In a runtime distributed context, one would prefer to gather information only about the recent past. This may be led by the need to save resources or by the fact that recent information is more relevant. In [78], we consider the *sliding window* model and propose two different (on-line) algorithms that approximate the items frequency in the active window. More precisely, we determine a (ε, δ)-additive-approximation meaning that the error is greater than ε only with probability δ. These solutions use a very small amount of memory with respect to the size N of the window and the number n of distinct items of the stream, namely, $O(\frac{1}{\varepsilon} \log \frac{1}{\delta} (\log N + \log n))$ and $O(\frac{1}{\tau\varepsilon} \log \frac{1}{\delta} (\log N + \log n))$ bits of space, where τ is a parameter limiting memory usage. We also provide their distributed variant, *i.e.*, considering the *sliding window functional monitoring* model, with a communication cost of $O(\frac{k}{\varepsilon^2} \log \frac{1}{\delta} \log N)$ bits per window (where k is the number of nodes). We compared the proposed algorithms to each other and also to the state of the art through extensive experiments on synthetic traces and real data sets that validate the robustness and accuracy of our algorithms.

Randomized Message-Passing Test-and-Set. In [101], 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 $p \leq n$ 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.

Throughput Prediction in Cellular Networks Downlink data rates can vary significantly in cellular net-works, with a potentially non-negligible effect on the user experience. Content providers address this problem by using different representations (*e.g.*, picture resolution, video resolution and rate) of the same content and switch among these based on measurements collected during the connection. If it were possible to know the achievable data rate before the connection establishment, content providers could choose the most appropriate representation from the very beginning. We have conducted a measurement campaign involving 60 users connected to a production network in France, to determine whether it is possible to predict the achievable data rate using measurements collected, before establishing the connection to the content provider, on the operator's network and on the mobile node. We show that it is indeed possible to exploit these measurements to predict, with a reasonable accuracy, the achievable data rate [81].

Population Protocol Model. The computational model of population protocols, introduced by Angluin and his colleagues in 2006, is a formalism that allows the analysis of properties emerging from simple and pairwise interactions among a very large number of anonymous finite-state agents. Significant work has been done so far to determine which problems are solvable in this model and at which cost in terms of states used by the protocols and time needed to converge. The problem tackled in [74] is the population proportion problem: each agent starts independently from each other in one of two states, say A or B, and the objective is for each agent to determine the proportion of agents that initially started in state A, assuming that each agent only uses a finite set of state, and does not know the number n of agents. We propose a solution which guarantees with any high probability that after $O(\log n)$ interactions any agent outputs with a precision given in advance, the proportion of agents that start in state A. The population proportion problem is a generalization of both the majority and counting problems, and thus our solution solves both problems. We show that our solution is optimal in time and space. Simulation results illustrate our theoretical analysis.

The context of [75] is the well studied dissemination of information in large scale distributed networks through pairwise interactions. This problem, originally called "rumor mongering", and then "rumor spreading", has mainly been investigated in the synchronous model. This model relies on the assumption that all the nodes of the network act in synchrony, that is, at each round of the protocol, each node is allowed to contact a random neighbor. In this paper, we drop this assumption under the argument that it is not realistic in large scale systems. We thus consider the asynchronous variant, where at time unit, a single node interacts with a randomly chosen neighbor. We perform a thorough study of T_n, the total number of interactions needed for all the n nodes of the network to discover the rumor. While most of the existing results involve huge constants that do not allow for comparing different protocols, we prove that in a complete graph of size $n \geq 2$, the probability that $T_n > k$ for all $k \geq 1$ is less than $(1 + 2k(n-2)^2/n)(1-2/n)^{k-1}$. We also study the behavior of the complementary distribution of T_n at point $cE(T_n)$ when n tends to infinity, in function of c. This paper received the Best Student Paper Award from the 15th IEEE Symposium on Network Computing and Applications (IEEE NCA 2016).

Bitcoin. Decentralized cryptocurrency systems offer a medium of exchange secured by cryptography, without the need of a centralized banking authority. Among others, Bitcoin is considered as the most mature one. Its popularity lies on the introduction of the concept of the blockchain, a public distributed ledger shared by all participants of the system. Double spending attacks and blockchain forks are two main issues in blockchain-based protocols. The first one refers to the ability of an adversary to use the very same bitcoin more than once, while blockchain forks cause transient inconsistencies in the blockchain. We show in [43], [89], [42] through probabilistic analysis that the reliability of recent solutions that exclusively rely on a particular type of Bitcoin actors, called miners, to guarantee the consistency of Bitcoin operations, drastically decreases with the size of the blockchain.

7.4. Future networks and architectures

Participants: Adlen Ksentini, Bruno Sericola, Yassine Hadjadj-Aoul, Jean-Michel Sanner, Hamza Ben Ammar.

SDN and NFV. Network Function Visualization (NFV) and Software Defined Network (SDN) currently play a key role to transform the network architecture from hardware-based to software-based.

SDN is in the process of revolutionizing the way of managing networks by providing a new way to support current and future services. However, by relocating the control functionality in a remote entity, the measurements' accuracy of the resources' utilization becomes more difficult, which complicates the decision making. Although there are previous works focusing on the problem of network management and measurement in SDN networks, only a few proposed solutions have taken into consideration the trade-off existing between statistics' polling frequency (i.e. generated overhead), and the accuracy of monitoring results (i.e. optimized resourcesâ allocation). In [62], we proposed a new approach calculating accurately the bandwidth utilization while adapting the polling frequency according to ports/switches activity. The emulations' results under Mininet clearly demonstrate the effectiveness of the proposed solution, which proved to be scalable compared to classical approaches. The controllers' placement is another important concern that emerged recently to solve the scalability and the reliability issues of SDN networks. The placement efficiency is influenced by both network operators (NO) strategy and the supported service requirements, which makes more complex the decision-making process. In particular, the need to support QoS-constrained services may lead NO to guide the controllers' placement in a way to ensure services efficiency while optimizing the underlying infrastructure. In [82] and [66], we proposed a model for the placement of network controllers, and we formulated a general optimization problem. To provide more flexibility and to avoid time-prohibitive calculations, we proposed a hierarchical clustering strategy for the controllers' placement allowing to minimize the number of network controllers while reducing the potential disparity of burden between the different controllers. Besides, the algorithms' structure makes it easy to act on other network parameters to improve the reliability of the SDN network. In [107], we proposed an improvement of such algorithms, by considering an evolutionary solution based on a genetic technique with an ad hoc cross-over operator designed to solve a mono-objective controller placement problem.

To connect the VNFs hosted in the same Data Center (DC) or across multiple DCs, virtual switches are required. Besides forwarding functions, virtual switches can be configured to mirror traffics for network management needs. Among the existing virtual switch solutions, Open vSwitch (OVS) is the most known and used. OVS is open source, and included in most of the existing Linux distributions. However, OVS performance in terms of throughput for smaller packets is very smaller than of line rate of the interface. To overcome this limitation, OVS was ported to Data Plane Development Kit (DPDK), namely OVDK. The latter achieves an impressive line rate throughput across physical interfaces. In [83], we presented the result of OVDK performance test when flow and port mirroring are activated, which was not tested so far. The performance test focuses on two parameters, throughput and latency in OVDK, allowing to validate the use of OVDK for flow forwarding and network management in the envisioned virtualized network architecture.

Mobile cloud. To cope with the tremendous growth in mobile data traffic on one hand, and the modest average revenue per user on the other hand, mobile operators have been exploring network virtualization and cloud computing technologies to build cost-efficient and elastic mobile networks and to have them offered as a cloud service. In such cloud-based mobile networks, ensuring service resilience is an important challenge to tackle. Indeed, high availability and service reliability are important requirements of carrier grade, but not necessarily intrinsic features of cloud computing. Building a system that requires the five nines reliability on a platform that may not always grant it is therefore a hurdle. Effectively, in carrier cloud, service resilience can be heavily impacted by a failure of any network function (NF) running on a virtual machine (VM). In [31], we introduce a framework, along with efficient and proactive restoration mechanisms, to ensure service resilience in carrier cloud. As restoration of a NF failure impacts a potential number of users, adequate network overload control mechanisms are also proposed. A mathematical model is developed to evaluate the performance of the proposed mechanisms. The obtained results are encouraging and demonstrate that the proposed mechanisms efficiently achieve their design goals.

Typically, maintaining a static pool of cloud resources to meet peak requirements with good service quality makes the cloud infrastructure costly. To cope with this, [58] proposes an approach that enables a cloud infrastructure to automatically and dynamically scale-up or scale-down resources of a virtualized environment aiming for efficient resource utilization and improved quality of experience (QoE) of the offered services. The QoE-aware approach ensures a truly elastic infrastructure, capable of handling sudden load surges while reducing resource and management costs. The paper also discusses the applicability of the proposed approach within the ETSI NFV MANO framework for cloud-based 5G mobile systems.

Video distribution. Due to the Internet usage evolution over these last years, the current IP-based architecture becomes heavier and less efficient for providing Internet services. In order to face this shortcoming, "Content Centric Networking" has been proposed. One of its important features is the use of in-network caching as a way of improving network performance and service scalability. However, in most of the existing CCN-based approaches several copies of the same content are present in the network, which reduce its efficiency. In [45], we proposed the "CLIque-based cooperative Caching" (CLIC) strategy, which basically consists in detecting cliques within the network topology to allocate more efficiently the content in the network. The main motivation of the proposed solution is to eliminate contents' redundancy between neighboring nodes while promoting the most popular contents. This approach guarantees a sufficient number of copies of popular files within the network while maximizing the number of distinct content items. We evaluate the proposed scheme through simulation. The results show significant improvements in terms of cache management and network performance.

In [59], we make the case for opening the telco CDN infrastructure to content providers by means of network function virtualization (NFV) and cloud technologies. We design and implement a CDN-as-a-Service architecture, where content providers can lease CDN resources on demand at regions where the ISP has presence. Using open northbound RESTful APIs, content providers can express performance requirements and demand specifications, which can be translated to an appropriate service placement on the underlying cloud substrate. To gain insight which can be applied to the design of such service placement mechanisms, we evaluate the capabilities of key enabling virtualization technologies by extensive testbed experiments.

Network design using new dependability metrics. When designing a network taking into account its capabilities face to possible failures to its components, the basic theoretical framework is classical network reliability, where the system under study is represented by a graph with perfect nodes and imperfect links randomly and independently failing. The corresponding connectivity-based metrics must then be evaluated in order to quantify the robustness of the networking architecture. Recently, a new family of metrics, called diameter-constrained, have been proposed and analyzed by Dionysos' collaborators and members. In [53], we developed some elements for a factoring theory associated with these metrics. The paper is focused on the detection of irrelevant components, a key task when evaluating these types of quantities using factorization. The paper also includes a factoring algorithm, which is an up-to-date procedure exploiting all available results for implementing the pivoting idea (proved to be one of the most powerful methods in classical reliability analysis).

In [54], we consider an homogeneous network (identical and independent components). In this context, if p is the probability that each of the components works, then any reliability metric is necessarily a polynomial in p, and computing these metrics can be reduced to counting problems (counting specific classes of paths or of cuts, for instance). In the paper, we quantify, in some sense, the "degree of difficulty" of these counting processes, and we identify the situations where they are "easy". The second contribution of the paper is to propose a fundamental problem from survivable network design, called the Network Utility Problem. The goal is to maximize network utility (defined as the opposite of the level of difficulty minus one), under a minimum edge-connectivity requirement.

Optical network design. Paper [65] presents a fast and accurate mathematical method to evaluate the blocking probability (the probability of a burst loss) in dynamic WDM networks without wavelength conversion (the present used technology). We assume that all links have the same number of wavelengths (the same capacity). The proposed model considers different traffic loads at each network connection (heterogeneous traffic). To take into account the wavelength continuity constraint, the method divides the network into a sequence of networks where all the links have capacity 1. Every network in the sequence is evaluated separately using an analytical technique. Then, a procedure combines the results of these evaluations in a way that captures the dependencies that occur in the real system due to the competition for bandwidth between the different connections. The method efficiently achieves results very close to those obtained by simulation, but orders of magnitude faster, allowing the evaluation of the blocking probability of all users (connections) for mesh network topologies.

7.5. 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 published in 2014 [109].

Network neutrality. Most of our activity has been devoted to the vivid network neutrality debate, going beyond the traditional for or against neutrality. We especially responded to the public consultation on draft BEREC Guidelines on implementation of net neutrality rules held during Summer 2016.

Network neutrality is often advocated by content providers, stressing that side payments to Internet Service Providers would hinder innovation. However, we also observe some content provider actually paying those fees. In [20] we intend to explain such behaviors through economic modeling, illustrating how side payments can be a way for an incumbent content provider to prevent new competitors from entering the market. We investigate the conditions under which the incumbent can benefit from such a barrier-to-entry, and the consequences of that strategic behavior on the other actors: content providers, users, and the Internet Service Provider. We also describe how the Nash bargaining solution concept can be used to determine the side payment.

In [105], we explain how non neutrality may be pushed by big CPs to their benefits. Major content/service providers are publishing grades they give to ISPs about the quality of delivery of their content. The goal is to inform customers about the "best" ISPs. But this could be an incentive for, or even a pressure on, ISPs to differentiate service and provide a better quality to those big content providers in order to be more attractive. This fits the network neutrality debate, but instead of the traditional vision of ISPs pressing content providers, we face here the opposite situation, still possibly at the expense of small content providers though. We design in [105] a model describing the various actors and their strategies, analyzes it thanks to non-cooperative game theory, and quantifies the impact of those advertised grades with respect to the situation where no grade is published. We illustrate that a non-neutral behavior, differentiating traffic, is not leading to a desirable situation.

While neutrality is focusing on the behavior of ISPs, we claim that the debate should be generalized. Indeed, the reality of the Internet in the 2010s is that various actors contribute to the delivery of data, with sometimes contradictory objectives. We highlight in [19] the fact that neutrality principles can be bypassed in many ways without violating the rules currently evoked in the debate. For example via Content Delivery Networks (CDNs), which deliver content on behalf of content providers for a fee, or via search engines, which can hinder competition and innovation by affecting the visibility and accessibility of content. We therefore call for an extension of the net neutrality debate to all the actors involved in the Internet delivery chain. We particularly challenge the definition of net neutrality as it is generally discussed. Our goal is to initiate a relevant debate for net neutrality in an increasingly complex Internet ecosystem, and to provide examples of possible neutrality rules for different levels of the delivery chain, this level separation being inspired by the OSI layer model.

The impact of a revenue-oriented CDN is particularly investigated in [104] and [70]. Content Delivery Networks (CDN) have become key telecommunication actors. They contribute to improve significantly the quality of services delivering content to end users. However, their impact on the ecosystem (end-users, the network operators and the content providers) raises concerns about their "neutrality", and therefore the question of their inclusion in the network neutrality debate becomes relevant. We compare the outcome with that of a neutral behavior, and at investigating whether some regulation should be introduced. We present a mathematical model and show that there exists a unique optimal revenue-maximizing policy for a CDN actor, in terms of dimensioning and allocation of its storage capacity, and depending on parameters such as prices for service/transport/storage. In addition, using the real traces, we compare the revenue-based policy with policies based on several fairness criteria. The CDN activity being potentially lucrative and not included in the neutrality debate, we analyze in [71] the revenue-optimal strategies and impact of a vertically integrated ISP-CDNs, which can sell those services to content providers. Our approach is based on an economic model of revenues and costs, and a multilevel game-theoretic formulation of the interactions among actors. Our model incorporates the possibility for the vertically-integrated ISP to partially offer CDN services to competitors in order to optimize the trade-off between CDN revenue (if fully offered) and competitive advantage on subscriptions at the ISP level (if not offered to competitors). Our results highlight two counterintuitive phenomena: an ISP may prefer an independent CDN over controlling (integrating) a CDN; and from the user point of view, vertical integration is preferable to an independent CDN or a no- CDN configuration. Hence, a regulator may want to elicit such CDN-ISP vertical integrations rather than prevent them.

Online platforms and search engines. Another set of key actors in the Internet economy is the online platforms and search engines. When a keyword-based search query is received by a search engine, a classified ads website, or an online retailer site, the platform has exponentially many choices in how to sort the search results. Two extreme rules are (a) to use a ranking based on estimated relevance only, which improves customer experience in the long run because of perceived quality, and (b) to use a ranking based only on the expected revenue to be generated immediately, which maximizes short-term revenue. Typically, these two objectives (and the corresponding rankings) differ. A key question then is what middle ground between them should be chosen.We introduce in [16] stochastic models that yield elegant solutions for this situation, and we propose effective solution methods to compute a ranking strategy that optimizes long-term revenues. This strategy has a very simple form and is easy to implement if the necessary data is available. It consists in ordering the output items by decreasing order of a score attributed to each. This score results from evaluating a simple function of the estimated relevance, the expected revenue of the link, and a real-valued parameter. We find the latter via simulation-based optimization, and its optimal value is related to the endogeneity of user activity in the platform as a function of the relevance offered to them.

The impact on other actors of search engines has led to the so-called search neutrality debate, as a parallel to the network neutrality debate. Search engines accused of biasing the ranking of their organic links to provide a competitive advantage to their own content. Based on the optimal ranking policy for a search engine obtained in [16], we investigate in [67] on an example whether non-neutrality impacts innovation. We illustrate that a revenue-oriented search engine may indeed deter innovation at the content level, hence the validity of the argument (without necessarily meaning that search engines should be regulated).

Sponsored auctions. Advertisement in dedicated webpage spaces or in search engines sponsored slots is usually sold using auctions, with a payment rule that is either per impression or per click. But advertisers can be both sensitive to being viewed (brand awareness effect) and being clicked (conversion into sales). In [33], [92], we generalize the auction mechanism by including both pricing components: the advertisers are charged when their ad is displayed, and pay an additional price if the ad is clicked. Applying the results for Vickrey-Clarke-Groves (VCG) auctions, we show how to compute payments to ensure incentive compatibility from advertisers as well as maximize the total value extracted from the advertisement slot(s). We provide tight upper bounds for the loss of efficiency due to applying only pay-per-click (or pay-per-view) pricing instead of our scheme. Those bounds depend on the joint distribution of advertisement visibility and population likelihood to click on ads, and can help identify situations where our mechanism yields significant improvements. We also describe how the commonly used generalized second price (GSP) auction can be extended to this context.

7.6. 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.

Rare event simulation. 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 (see our book [108] for a global introduction to the field). This is the main problem in the area. Dionysos' work focuses then on dealing with the rare event situation, with a particular focus on dependability [40].

A non-negligible part of our activity on the application of rare event simulation was about the evaluation of static network reliability models. In a static network reliability model 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 [12], 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 zero, whereas GS does not have this property. But when the size of the network (or the number of shocks) increases, PMC and turnip eventually fail, whereas GS works nicely (empirically) for very large networks, with over 5000 shocks in our examples. In [73], we propose a methodology for calibrating a dependent failure model to compute the reliability in a telecommunication network, following a similar starting point (that is, using Marshall-Olkin copulas). In practice, this model is difficult to calibrate because it requires the estimation of a number of parameters that is exponential in the number of links. We formulate an optimization problem for calibrating a Marshall-Olkin copula model to attain given marginal failure probabilities for all links and the correlations between them. Using a geographic failure model, we calibrate various Marshall-Olkin copula models using our methodology, we simulate them, and we benchmark the reliabilities thus obtained. Our experiments show that considering the simultaneous failures of small and connected subsets of links is the key for obtaining a good approximation of reliability, confirming what it is suggested by the telecommunication literature.

A related problem is when links have random capacities and a certain target amount of flow must be carried from some source nodes to some destination nodes is considered in [47]. Each destination node has a fixed demand that must be satisfied and each source node has a given supply. The goal is to estimate the unreliability of the network, defined as the probability that given the realized link capacities, the network cannot carry the required amount of flow to meet the demand at all destination nodes. We adapt GS and PMC to this context. In [55], we explore other methods designed to reduce the variance of the estimators in this context. All of them are adaptations of methods originally developed to make reliability estimations on different network models. These methods are introduced together with a brief review of the algorithms on which they are based.

A new application of our previously designed zero-variance approximation importance sampling method has been developed in [76]: To accurately estimate the reliability of highly reliable rail systems and comply with contractual obligations, rail system suppliers such as ALSTOM require efficient reliability estimation techniques. While in our previous works, the studied graph models were dealing with failing links, we propose an adaptation of the algorithm to evaluate the reliability of real transport systems where nodes are the failing components. This is more representative of railway telecommunication system behavior. Robustness measures of the accuracy of the estimates, bounded or vanishing relative error properties, are discussed and results from a real network (Data Communication System used in automated train control system) showing bounded relative error property, are presented.

Random variable generation. Simulation requires the use of pseudo-random generators. In [18], we examine the requirements and the available methods and software to provide (or imitate) uniform random numbers in parallel computing environments. In this context, for the great majority of applications, independent streams of random numbers are required, each being computed on a single processing element at a time. Sometimes, thousands or even millions of such streams are needed. We explain how they can be produced and managed. We devote particular attention to multiple streams for GPU devices.

Sampling from the Normal distribution truncated to some finite or semi-infinite interval is of particular interest for certain applications in Bayesian statistics, such as to perform exact posterior simulations for parameter inference. We study and compare in [46] various methods to generate such random variables, with special attention to the situation where the interval is far in the tail. The algorithms are implemented and available in Java, R, and MATLAB, and the software is freely available.

Quasi-Monte Carlo (QMC). Finally, we have continued our work on QMC methods. In [15], we review the Array-RQMC method, its variants, sorting strategies, and convergence results. We are interested in the convergence rate of measures of discrepancy of the states at a given step of the chain, as a function of the sample size, and also the convergence rate of the variance of the sample average of a (cost) function of the state at a given step, viewed as an estimator of the expected cost. We summarize known convergence rate results and show empirical results that suggest much better convergence rates than those that are proved. We also compare different types of multivariate sorts to match the chains with the RQMC points, including a sorting procedure based on a Hilbert curve.

The description of a new software tool and library named Lattice Builder, written in C++, that implements a variety of construction algorithms for good rank-1 lattice rules (a familiy of sequences used in QMC methods) is provided in [17]. The library is extensible, thanks to the decomposition of the algorithms into decoupled components, which makes it easy to implement new types of weights, new search domains, new figures of merit, etc.

7.7. Wireless Networks

Participants: Osama Arouk, Btissam Er-Rahmadi, Adlen Ksentini, Meriem Bouzouita, Pantelis Frangoudis, Yassine Hadjadj-Aoul, César Viho, Quang Pham, Gerardo Rubino.

We are continuing our activities around wireless and mobile networks, by focusing more on leveraging the current mobile and wireless architecture toward building the 5G systems.

Congestion control for M2M applications. Machine-to-Machine (M2M) communications are expected to be one of the major drivers for the future 5G network. It is expected that M2M will come up with substantial revenue growth for Mobile Network Operators (MNO), but they represent at the same time the most important challenge they are facing. For instance, a massive number of Machine-to-Machine (M2M) devices performs simultaneously Random Accesses (RA), which causes severe congestions and reduces the RA success probability. To control the Radio Access Network (RAN) overload and alleviate the congestion between M2M devices, 3GPP developed the Access Class Barring (ACB) procedure that depends on an access probability called the ACB factor. In [48][24], we first presented a simple fluid model of M2M devices' random access. This model is then used to derive an optimal regulator of the ACB factor based on nonlinear non affine control theory. The main advantages of the proposed approach are twofold. First, the proposal is fully compliant with the standard while it reduces significantly the computation and the signaling overheads. Second, it provides an efficient mean to regulate adaptively the ACB factor as it guarantees having an optimal number of M2M devices accessing concurrently to the RAN. The obtained results based on simulations show clearly the robustness of the proposed approach, and its superiority compared to existing proposals. However, such a model assumes a perfect knowledge about the number of M2M attempting the ACB and the RA, which is not possible in realistic use cases. For this reason, we proposed in [50] a system-agnostic controller, which computes the barring factor dynamically based only on the mismatch between the average number of M2M devices succedding in the RA and the optimal number of M2M which should succeed. We base our controlling algorithm in a Propostional Integral Derivative (PID)-based controller. Simulation results show that the algorithm outperforms the existing solutions by improving significantly the access success probability while minimizing radio resources' underutilization.

Different schemes were proposed in the literature to solve the congestion problem by regulating the M2M devices' opportunities of transmission. Nonetheless, as revealed in [51], these schemes turn out to be ineffective in case of heavily congested M2M networks. In fact, in such a condition, the unpredictable and increasingly accumulated number of devices cannot be blocked. This augments the risk of M2M devices' synchronized access, which may result in a congestion collapse. Consequently, we proposed, in [49], a methodology for a better estimation of the number of M2M devices attempting the access. We also proposed a novel implementation of the ACB process, which dynamically computes the ACB factor according to the network's overload conditions and includes a corrective action adapting the controller work, based on the mismatch existing between the computed and the targeted mean load. The simulation results show that the proposed algorithms allow improving considerably the estimation of the number of M2M devices' arrivals, while outperforming existing techniques.

In [32], we proposed a novel approach to deal with massive synchronous access attempts, tailored for both M2M delay-sensitive applications and energy constrained ones. The main idea behind the paper is to leverage crowd sourcing data, transmitted from the devices succeeding in the RACH procedure, to tune the access parameters, without requiring too complex techniques for the estimation of the number of attempts. Simulation results show that the proposed scheme achieves sub-optimal performance in the wireless resources' utilization while reducing significantly both the number of access attempts and the access latency for delay sensitive applications. This allows guaranteeing energy conservation.

In [44], we proposed two optimal solutions that use Device-to-Device (D2D) communications to lightweight the overhead of M2M devices on 5G networks. Each scheme has a specific objective, and aims to manage the communications between devices and eNodeBs to achieve its objective. The proposed solutions nominate the devices that should communicate using D2D communications and those that should directly communicate with eNodeBs. The first solution aims to reduce the energy consumption, whereas the second one aims to reduce the data transfer delay at the eNodeBs. The performance of the proposed schemes is evaluated via simulations and the obtained results demonstrate their feasibility and ability in achieving their design goals.

Network selection and optimization. With the explosion of mobile data traffic, the Fixed and Mobile Converged (FMC) network are being heavily required. Mobile devices have the capability of connecting simultaneously to different access networks in the FMC architecture. Access network selection becomes an issue when mobile devices are under coverage of different access networks, since a bad selection may lead to network congestion and degrade the QoE of users. In order to address this problem, in [91] we modeled and analyzed the interface selection procedure using control theory. Based on our model, we designed a controller which can send to mobile devices a network selection command calculated instantaneously for the access network selection.

Dynamic Adaptive Streaming over HTTP (DASH), with its different proprietary versions, is presently the most widely adopted technology for video delivery over the Internet. DASH offers significant advantages, enabling users to switch dynamically between different available video qualities responding to variations in the current network conditions during video playback. This is particularly interesting in wireless and mobile access networks, which present such variations in a hard to predict manner, but sometimes quite frequently. Moreover, mobile users of these networks share a common radio access link and, thus, a common bottleneck in case of congestion, which may cause user experience to degrade. In this context, the Mobile Edge Computing (MEC) emerging standard gives new opportunities to improve DASH performance, by moving IT and cloud computing capabilities down to the edge of the mobile network. In [69] and [103] we proposed a novel architecture for adaptive HTTP video streaming tailored to a MEC environment. The proposed architecture includes an adaptation algorithm running as a MEC service, aiming to relax network congestion while improving the Quality of Experience (QoE) for mobile users. Our mechanism is standards-compliant and compatible with receiver-driven adaptive video delivery algorithms, with which it cooperates in a transparent manner.

Low-rate wireless personal area networks (WPANs) (and also wireless sensor networks) suffer from many constraints. The IEEE 802.15.4 standard proposes the slotted CSMA/CA as a communication channel access mechanism with collision avoidance that takes into account the constraints of WPANs. In [22], we proposed to introduce a data fragmentation mechanism into slotted CSMA/CA to improve a bandwidth utilisation. The novelty here is the use of the fragmentation mechanism to replace an acknowledgement frame after the transmission of the fragment and the remaining frame. The beacon frame is used to confirm the success transmission of a data fragment. To evaluate the performance of our proposition, we have developped a three dimension Markov chain which modelises the behaviour of the node using IEEE 802.15.4 with data fragmentation mechanism without using an ACK frame. The analytical results concerning the network throughput and the transmission success delay demonstrate the improvement of the bandwidth occupation.

Mobile networks' improvements. In [85], we introduced the concept of elastic bearer in Evolved Packet System (EPS), which allows the users to enhance on-demand the performance of certain applications and permits the network to efficiently manage the resource allocation taking into account the application type. In particular, the paper introduces a set of mechanisms to trigger and support bearer elasticity in EPS based on the Quality of Experience (QoE) perceived by users or based on feedback from Radio Access Network (RAN). Bearer elasticity can be attained through potential Packet Data Network/Serving Gateway (PDN/S-GW) relocation to eventually improve QoE within and beyond the mobile network operator premises. The paper also introduces a set of methods to identify and cope with a storm of requests for particular applications at densely populated areas.

One important objective of 5G mobile networks is to accommodate a diverse and ever-increasing number of user equipment (UEs). Coping with the massive signaling overhead expected from UEs is an important hurdle to tackle so as to achieve this objective. In [11], we devised an efficient tracking area list management (ETAM) framework that aims for finding optimal distributions of tracking areas (TAs) in the form of TA lists (TALs) and assigning them to UEs, with the objective of minimizing two conflicting metrics, namely paging overhead and tracking area update (TAU) overhead. ETAM incorporates an online part and an and offline one, in order to achieve its design goal. In the online part, two strategies were proposed to assign in real time, TALs to different UEs, while in the offline part, three solutions were proposed to optimally organize TAs into TALs. The performance of ETAM is evaluated via analysis and simulations, and the obtained results demonstrate its feasibility and ability in achieving its design goals, improving the network performance by minimizing the cost associated with paging and TAU.

QoE aware routing in wireless networks. This year we continued our research on QoE-based optimization routing for wireless mesh networks. First, we approximate PSQA models by explicit mathematical forms, which can be used to find the optimal or near to optimal routes. Next, the hardness of the problem is studied and decentralized algorithms are proposed. The quality of the solution, computational complexity of the proposed algorithm, and the fairness are the main concerns of this work. Several centralized approximation algorithms have been proposed in order to address the complexity and the quality of the published solutions. The results can be found in the following papers: [25],[94], [95] and [26]. However, these centralized algorithms are not appropriate in large-scale networks. Thus, a distributed algorithm is necessary as a complement of the existing centralized methods. This is currently studied at the team.

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Contract with Industry: ALSTOM-Inria Common Lab

Participants: Bruno Tuffin, Gerardo Rubino.

Bruno Tuffin is the co-director of ALSTOM-Inria common Lab.

Dionysos manages a project with ALSTOM on system availability simulation taking into account logistic constraints. Current ALSTOM Transport and Power contracts, especially service-level agreements, impose stringent system availability objectives. Non-adherence to the required performance levels often leads to penalties, and it is therefore critical to assess the corresponding risk already at a tender stage. The challenge is to achieve accurate results in a reasonable amount of time. Monte Carlo simulation provides estimates of the quantities it is desired to predict (e.g., availability). Since we deal with rare events, variance reduction techniques, specifically Importance Sampling (IS) here, is used. The goal of the project is to establish the feasibility of IS for solving problems relevant to ALSTOM and to develop the corresponding mathematical tools.

8.2. Bilateral Contract with Industry: Participation in a CRE with Orange

Participant: Bruno Tuffin.

We are participating to a CRE (managed by Telecom Bretagne) with Orange on the strategies of Content Delivery Networks (CDNs) and their impact on the overall Internet economy and regulation. In this study, we focus on the CDN as an economic actor. The goals are 1) to analyze CDNs' caching strategies from an economic point of view, 2) to study the strategies of an integrated CDN actor, and 3) to study the impact of CDNs in the net neutrality debate.

8.3. Cifre contract on Device-Assisted Distributed Machine-Learning on Many Cores

Participants: Corentin Hardy, Bruno Sericola.

This is a Cifre contract including a PhD thesis supervision (PhD of Corentin Hardy), done with Technicolor. The starting point of this thesis is to consider the possibility to deploy machine-learning algorithms over many cores, but out of the datacenter: on the devices (home-gateways) deployed by Technicolor in users' homes. In this device-assisted view, an initial processing step in the device may significantly reduce the burden on the datacenter back-end. Problems are numerous (power consumption, CPU power, network bandwidth and latency), but costs for the operator can be lowered and scale may bring some new level in data processing.

8.4. Cifre contract on SDN for 5G mobile networks

Participants: César Viho, Yassine Hadjadj-Aoul, Adlen Ksentini.

This is a Cifre contract (2015-2018) including a PhD thesis supervision (PhD of Imad Alawe), done with TDF, on cooperation in SDN use for the 5th generation of mobile networks. The objective of the thesis is to study and devise appropriate solutions to introduce SDN in the current LTE architecture.

8.5. Camion

Participants: Yassine Hadjadj-Aoul, César Viho, Raymond Marie.

We are working in the 2-year (October 2014 to October 2016) Eurostars European Project Camion, which aims at offering cost-efficient, QoE-optimized content delivery, allowing for faster content access, as well as offline operation, while improving wireless network capacity and coverage. Camion is leaded by JCP-Connect, and the partners are a SME (FON) and our team. The project is extended until June 2017.

8.6. DVD2C

Participants: Yassine Hadjadj-Aoul, Adlen Ksentini, Pantelis Frangoudis.

We are working in the 3-year (September 2014 – September 2017) FUI Project DVD2C, which aims to virtualize CDN through the Cloud and Network Function Virtualization concept. DVD2C is leaded by Orange labs., and the partners are two SMEs (Viotech and Resonate) and two academics (our team and Télécom Paris Sud).

9. Partnerships and Cooperations

9.1. National Initiatives

9.1.1. ANR

- Adlen Ksentini is participating at 20% of his time to the IRT BCOM granted by the ANR.
- Yassine Hadjadj-Aoul is participating at 20% of his time to the IRT BCOM granted by the ANR.
- Yann Busnel is a member of the two following projects: Inhare granted by the ANR (ANR-15-CE19-0024) and DeSCeNt granted by the LabEx CominLabs (ANR-10-LABX-07-01).

9.2. Inria Project Labs

Participants: Yassine Hadjadj-Aoul, Gerardo Rubino, Bruno Tuffin.

Inria Project Labs' (IPL) initiatives enable the launch of ambitious research projects directly linked with the institute.

9.2.1. BetterNet

BetterNet aims at building and delivering a scientific and technical collaborative observatory to measure and improve the Internet service access as perceived by users. In this Inria Project Lab, we will propose new original user-centered measurement methods, which will associate social sciences to better understand Internet usage and the quality of services and networks. Our observatory can be defined as a vantage point, where:

- tools, models and algorithms/heuristics will be provided to collect data,
- acquired data will be analyzed, and shared appropriately with scientists, stakeholders and civil society,
- and new value-added services will be proposed to end-users.

Inria Project Teams involved: Diana, Dionysos, Inria Chile, Madynes, Muse, Spirals

9.3. European Initiatives

9.3.1. FINTEROP

Program: H2020-ICT-12-2015

Project acronym: F-Interop

Project title: FIRE+ online interoperability and performance test tools to support emerging technologies from research to standardization and market launch

Duration: November 2015 – October 2018

Coordinator: UPMC-LIP6

Other partners: 9 partners including our team Dionysos (F. Sismondi and C. Viho), and Eva (T. Watteyne)

Abstract: The goal of F-Interop is to extend FIRE+ with online interoperability and performance test tools supporting emerging IoT-related technologies from research to standardization and to market launch for the benefit of researchers, product development by SME, and standardization processes.

9.3.2. Collaborations with Major European Organizations

Partner 1: Sapienza University of Rome, Italy.

We work with Nicoló Rivetti and Leonardo Querzoni on the analysis of stream processing systems.

9.4. International Initiatives

9.4.1. Inria International Partners

9.4.1.1. MOCQUASIN

Title: Monte Carlo and Quasi- Monte Carlo for rare event simulation

International Partner (Institution - Laboratory - Researcher):

University of Montreal (Canada)

Duration: started in 2013

See also: http://www.irisa.fr/dionysos/pages_perso/tuffin/MOCQUASIN/

The goal of this team is to compute integrals, sums or to solve equations or optimization problems by means of Monte Carlo methods, which are statistical tools used when the models have a high complexity (for instance a large dimension). They are unavoidable methods in areas such as finance, electronics, seismology, computer science, engineering, physics, transport, biology, social sciences... Nonetheless, they have the reputation of being slow, i.e. to require a large computational time to reach a given precision. The goal of the project is to work on acceleration techniques, meaning methods allowing to reach the targeted precision in a shorter computational time than with the standard procedure. A typical framework is that of rare event simulation for which getting even only one occurrence of the event could require a too long computing time. In this case, there are two main acceleration techniques: importance sampling and splitting, on which we work.

9.4.1.2. Collaborations with the UTFSM at Valparaíso, Chile

We maintain a strong line of collaborations with the Technical University Federico Santa María (UTFSM), Valparaíso, Chile. Over the years, this has taken different forms (associated team Manap, Stic AmSud project "AMMA", Stic AmSud project "DAT", see next module). Currently, we have a joint PhD work running (PhD of Nicolás Jara, to be defended in 2017), and a new joint PhD to be started in 2017 (PhD of Jonathan Olavarría). The first one is on optical network analysis and design, the second one on modeling evaluation techniques.

9.4.1.3. International Initiatives

DAT

Title: Dependability Analysis Tool

International Partners:

Prof. H. Cancela, Univ. of the Republic, Computer Science, Uruguay;

Prof. R. Vallejos, UTFSM, Valparaíso, Electrical Eng., Chile;

G. Rubino, Dionysos, Inria, general responsible for the project.

Duration: 2015 - 2016

Start year: 2015

The main scientific objective of this project is to develop new techniques to assess the most important dependability properties of a complex system subject to the failures and possible repairs of its components. The central argument behind our proposal is our previous work in the area and some unpublished preliminary and promising results that we believe deserve deep exploration and that should lead to faster evaluation procedures than those available today. We also intend to implement these techniques in an integrated software package usable both in industry and for teaching purposes. Concerning applications, again based on the skills of the participating teams and our past common work, we will illustrate our findings on problems coming from the wireless and optical networking domains.

SM-HCD-HDD

Title: Statistical methods for highly complex and/or high dimensional data

International Partners:

Prof. Ricardo Fraiman, Mathematics, Univ. of the Republic, Uruguay, head of the project;

many partners in Uruguay, Argentina and Chile,

G. Rubino for Dionysos, Inria

Duration: 2016 - 2017

Start year: 2016

The aim of this project is to develop theoretical and computational tools to solve statistical problems that occur when data leaves in high dimensional spaces and/or lives in complex spaces that induce complex geometries.

9.5. International Research Visitors

9.5.1. Visits of International Scientists

Jebali Ameni, from INSAT (Tunisia)

Jorge Graneri, from UDELAR, Uruguay

Héctor Cancela, from UDELAR, Uruguay

Franco Robledo, from UDELAR, Uruguay

Claudio Risso, from UDELAR, Uruguay

Reinaldo Vallejos, from UTFSM, Chile

Marta Barría, from university of Valparaíso, Chile

9.5.1.1. Research Stays Abroad

Yann Busnel has been granted as an invited professor at La Sapienza Universitá di Roma, Italy, for 3 months from March to June 2016.

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events Organisation

10.1.1.1. Member of the Organizing Committees

Pierre L'Ecuyer is member of the Steering Committee of MCQMC.

G. Rubino and B. Tuffin are members of the Steering Committee of the International Workshop on Rare Event Simulation (RESIM).

Y. Hadjadj-Aoul is a member of the steering committee of the International Conference on Information and Communication Technologies for Disaster Management (ICT-DM) from December 2015.

Yann Busnel has been "Habilité à diriger les recherches" at the École Normale Supérieure de Rennes, under the seal of University Bretagne Loire, in December 2016, presenting a defense entitled "Analyse et traitement de flux de données large échelle".

10.1.2. Scientific Events Selection

10.1.2.1. Chair of Conference Program Committees

- Bruno Tuffin served as TPC co-chair for the 10th International Conference Performance Evaluation Methodologies and Tools (Valuetools'16), September 2016, Taormina, Italy.
- Yassine Hadjadj-Aoul served as TPC co-chair for the 3rd International Symposium on Networks, Computers and Communications (ISNCC 2016), May 2016, Hammamet, Tunisia.
- Yassine Hadjadj-Aoul served as TPC co-chair for the 13th conférence francophone sur les nouvelles technologies de la répartition (NOTERE 2016), Paris, France.

10.1.2.2. Member of the Conference Program Committees

Pierre L'Ecuyer was a member of the program committee of the following events:

- MCQMC'2016: Twelve International Conference on Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing, Stanford, California, Aug. 2016.
- SIMULTECH 2016: International Conference on Simulation and Modeling Methodologies, Technologies and Applications, Lisbon, Portugal, July 2016.
- ICORES 2016: International Conference on Operations Research and Enterprise Systems, Rome, Italy, Feb. 2016.

Bruno Tuffin was a member of the program committee of the following events:

- 5th Workshop on Smart Data Pricing (SDP 2016), Workshop of IEEE INFOCOM 2016, San Francisco, USA, April 2016.
- IEEE ICC 2016 - Communications Software, Services and Multimedia Applications Symposium, 23-27 May 2016, Kuala Lumpur, Malaysia.
- 22nd Asia-Pacific Conference on Communications (APCC 2016), Yogyakarta, Indonesia on 25-27 August 2016.
- GECON2016 (13th International Conference on Economics of Grids, Clouds, Systems, and Services), 20-22 September 2016, Athens, Greece.
- The International Conference on Wireless Networks and Mobile Communications (WINCOM?16). Fez, Morocco, October 26-29, 2016.
- Globecom'16 - CSSMA (2016 IEEE Global Communications Conference: Communications Software, Services and Multimedia Apps), 4-8 December 2016, Washington VD, USA.

Bruno Sericola was member of the program committee of the following events:

- ASMTA 2016: 23rd International Conference on Analytical and Stochastic Modelling Techniques and Applications, Cardiff, Wales, UK, August 2016.
- MACOM 2016: 9th International Workshop on Multiple Access Communications, Aalborg, Denmark, November 2016.
- INTERNET 2016: 8th International Conference on Evolving Internet, Barcelona, Spain, November 2016.
- DEPEND 2016: 9th International Conference on Dependability, Nice, France, July 2016.

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

- NCA 2016: 15th IEEE International Symposium on Network Computing and Applications, Boston, USA, October 2016.
- CoRes 2016: 1ères Rencontres Francophones sur la Conception de Protocoles, l'Évaluation de Performance et l'Expérimentation des Réseaux de Communication, Bayonne, France, May 2016.

Adlen Ksentini was member of the program committee of the following events:

- IEEE ICC 2016 Mobile Wireless Networks Symposium, Kuala Lumpur, Malaysia 2016.
- IEEE Globecom 2016 Mobile Wireless Networks Symposium, Washington, DC USA 2016.
- IEEE WCNC 2016, Doha, Qatar, 2016.

Yassine Hadjadj-Aoul was a member of the program committee of the following events:

- IEEE ICC 2016 Mobile Wireless Networks Symposium, Kuala Lumpur, Malaysia 2016.
- IEEE Globecom 2016 Mobile Wireless Networks Symposium, Washington, DC USA 2016.
- IEEE WCNC 2016, Doha, Qatar, 2016.
- IEEE CSCN 2016 IEEE Conference on Standards for Communications & Networking, Berlin, Germany, 2016.

Gerardo Rubino was a member of the program committee of the following events:

- Networking 2016, The IFIP Networking Conference, Vienne, Autriche, May 2016.
- MAM9 (9th) International Conference on Matrix Analytic Methods in Stochastic Models, Budapest, Hongrie, June 2016.
- COMTEL 2016, VIII International Congress in Computer Science and Telecommunications, Lima, Perú, September 2016.
- Mascots 2016, the IEEE International Symposium on Modelling, Analysis and Simulation of Computer and Telecommunications Systems, Londres, UK, September 2016.

10.1.3. Journal

10.1.3.1. Member of the Editorial Boards

Bruno Tuffin is the Simulation Area Editor for *Informs Journal on Computing*.

Bruno Tuffin is an associate editor for the following journal:

- Mathematical Methods of Operations Research, since November 2008.

Pierre L'Ecuyer is an associate editor for the following journals:

- ACM Transactions on Mathematical Software, since August 2004.
- Statistics and Computing (Springer-Verlag), since June 2003.
- International Transactions in Operational Research, since May 2007.

Bruno Sericola is an associate editor for the following journals:

- International Journal of Stochastic Analysis, since April 2010.
- Performance Evaluation, since April 2015.

Bruno Sericola is Editor in Chief of the books series “Stochatic Models in Computer Science and Telecommunications Networks”, ISTE/WILEY, since March 2015.

Yassine Hadjadj-Aoul is a member of the editorial broad of the CSC “International Journal of Computer Networks (IJCN)” since 2012.

Yassine Hadjadj-Aoul was a member of the guest editorial board of the following journals:

- Special issue in IGI, International Journal of Distributed Systems and Technologies (IJDST), “Information and Communication Technologies for Disaster Management” [99], Vol. 7, Issue 4, ISSN: 1947-3532 (October 2016)
- Hindawi Mobile Information Systems (Editorial), “Wireless and Mobile Technologies for the Internet of Things” [100], vol. 2016, ID 8206548 (2016)

10.1.3.2. Reviewer - Reviewing Activities

In addition to the reports done during his associate editor and conference TPC member duties, Bruno Tuffin has reviewed papers in 2016 for COMNET, Operations Research, Stochastic Models, Computers & OR, Journal of Communications & Networks.

Bruno Sericola served as a reviewer for several major international journals and conferences.

Adlen Ksentini has reviewed papers in 2016 for Transaction Paralel Distributed Systems-TPDS, Transaction on Wireless Communication - TWC, Transaction on Vehicular Technology - TVT, Wireless Communication Magazine, Elsevier Computer Communication.

Yassine Hadjadj-Aoul has reviewed papers in 2016 for Transaction on Vehicular Technology-TVT, Wireless Communication Magazine, Elsevier Computer Communication and for the following international conferences: Globecom'16, ICC'16, WCNC'16 and others.

César Viho has reviewed project proposals for the ANR and for CIFRE contracts for the ANRT. He has reviewed papers for the journals IEEE Transaction on Wireless Communication, IEEE Transactions on Vehicular Communications, IEEE Communications Magazine, and for the following international conferences: IWCNC, Globecom, and CCNC.

Gerardo Rubino has reviewed papers for several international journals and conferences (Networking 2016, MAM9, Mascots 2016, ...), ANR projects, etc.

10.1.4. Invited Talks

B. Tuffin. gave a plenary talk "Some applications of importance sampling to dependability analysis" at the International Conference on Monte Carlo techniques, Paris, France, July 2016.

B. Tuffin gave the following seminar presentations:

- B. Tuffin. Net neutrality. IRISA Evaluation by HCRES, Jan. 2016.
- P. L'Ecuyer, P. Maillé, N. Stier-Moses, B. Tuffin. Stochastic Optimization for Search Engines: Revenue-Maximizing Rankings with Quality-Sensitive Consumers. CREM-IRISA Workshop on Network Economics, January 2016.
- P. Maillé, G. Simon, B. Tuffin. On revenue-oriented Content Delivery Networks and their impact on Net Neutrality. CREM-IRISA Workshop on Network Economics, January 2016.
- B. Tuffin. La neutralité du Net. Journées scientifiques Inria, Rennes, Juin 2016.
- P. Maille and B. Tuffin. Auctions For Online Ad Space Among Advertisers Mixing Pay-per-view And Pay-per-click. DIRO, Université de Montréal, Canada, Nov 30, 2016.

Y. Hadjadj-Aoul gave a plenary talk "Efficient support of Massive M2M communications during emergency situations" at the International Conference on Information and Communication Technologies for Disaster Management (ICT-DM 2016), Vienna, Austria, December 2016.

Y. Busnel gave a plenary talk "Comment générer des échantillons uniformes sur des grandes masses de données ?" at the 9ème Colloque Francophone sur les sondages, Gatineau, Canada, October 2016.

G. Rubino made several invited and keynote talks in 2016:

- an invited talk at Advances in Uncertainty Quantification Methods, Algorithms and Applications, Kaust, Saudi Arabia, about rare event analysis [37];
- an invited talk at SIAM Conference on Uncertainty Quantification, Lausanne, Suisse, about rare event analysis in static models [39];
- a keynote at Mascots'16, London, UK, about a new performance metric [34];
- an invited talk at Technoconférence "Nouvelles expériences son et vidéo", Rennes, France, about QoE measuring [35];
- a keynote in Math AmSud Seminar, Montevideo, Uruguay, about QoE analysis and Big Data problems [38];
- a keynote at SoCPaR'16, Vellore, India, about new results concerning Perceptual Quality measuring [36].

10.1.5. Leadership within the Scientific Community

Bruno Tuffin is the co-director of the common lab ALSTOM-Inria since 2014.

Gerardo Rubino is one of the three French representatives at the Scientific Committee of the IFCAM (Indo-French Centre for Applied Mathematics), managing the cooperation in mathematics of the two countries, and federating at the French side, among several other participants, Inria and CNRS.

Gerardo Rubino is a member of the CSV (the technical committee) of the Images and Networks Cluster of Brittany, France.

10.1.6. Research Administration

- Bruno Sericola is responsible for the Inria Rennes-Bretagne Atlantique budget.
- Bruno Sericola is the leader of the research group MAPI (Math Appli Pour l'Info) the goal of which is to improve the collaboration between computer scientists and mathematicians.

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

Master 2nd year: Bruno Sericola, Mathematics, 12 hours, M2, Istic/University of Rennes 1, France.

Master 2nd year: Bruno Sericola, Logistic and performance, 12 hours, M2, Faculté de sciences économiques/University of Rennes 1, France.

Master 2nd year: Gerardo Rubino, Performance Evaluation, Istic/University of Rennes 1, France, 16 hours.

Master 2nd year: Performance Evaluation, Bruno Tuffin, Istic/University of Rennes 1, France, 4 hours.

Master 2nd year: César Viho, Algorithms on graphs, 40 hours, Istic/University of Rennes 1, France

Master 2nd year: César Viho, Multimedia Networking, 8 hours, Istic/University of Rennes 1, France

Master 2nd year: Yassine Hadjadj-Aoul, Multimedia streaming over IP (MMR), 48 hours, Esir/University of Rennes 1, France

Master 2nd year: Yassine Hadjadj-Aoul, Multimedia services in IP networks (RSM), 29 hours, Esir/University of Rennes 1, France

Master 2nd year: Yassine Hadjadj-Aoul, Software Defined Networks, 4 hours, Istic/University of Rennes 1, France

Master 2nd year: Yassine Hadjadj-Aoul, Video streaming over IP, 8 hours, Istic/University of Rennes 1, France

Master pro. 2nd year: Yassine Hadjadj-Aoul, C Programming, 12 hours, Istic/University of Rennes 1, France

ENSAI Rennes 3rd year: Gerardo Rubino, Neural Networks, 9 hours.

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

Seminars: Yann Busnel, Big Data for Security and Safety, 12 hours, Sapienza Universitá di Roma, Italy, April 2016.

Seminars: Yann Busnel, Big Data for Security and Safety, 18 hours, Indian Institute of Technology, Indore, India, October 2016.

10.2.2. Supervision

PhD in progress: Ajit Rai, "Availability prediction with logistics", started in May 2015; advisors: B. Tuffin & G. Rubino, University Rennes 1.

PhD in progress: Corentin Hardy,"Device-Assisted Distributed Machine-Learning on Many Cores", started in April 2016; advisors: Bruno Sericola & Erwan Le Merrer from team Cidre, University Rennes 1.

PhD in progress: Yves Mocquard, “Analyse de flux de données massifs dans les systèmes distribués large échelle”, started on September 2015; advisors: Bruno Sericola and Emmanuelle Anceaume from team Cidre, University Rennes 1.

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

PhD in progress: Hamza Ben Ammar, “Socially-aware network and cache resources optimization for efficient media content delivery in Content Centric Networks”, started in October 2015; advisors: Yassine Hadjadj-Aoul, Adlen Ksentini and Soraya Ait Chellouche, University Rennes 1.

PhD in progress: Imad Alawe, “Mobile SDN architecture”, started in October 2015; advisors: César Viho, Yassine Hadjadj-Aoul, University Rennes 1, Philippe Bertin, B-COM and Davy Darche, TDF.

PhD in progress: Jean-Michel Sanner; Cifre Grant, Orange Labs, “SDN technologies for network services performances adaptation of carriers networks”; started on January 2013; Advisors: Y. Hadjadj-Aoul and G. Rubino; University Rennes 1.

PhD in progress: Yue Li; Cifre Grant, Orange Labs; title: “Elaboration d’une architecture réseau unifiée, ouverte et flexible”, started on October 2013; Advisors: Y. Hadjadj-Aoul and G. Rubino; University Rennes 1.

PhD in progress: Nicolás Jara, “Fault tolerant design of dynamic WDM optical networks”, Technical University Federico Santa María (UTFSM) and university of Rennes 1, France. Advisors: R. Vallejos (Chile) and G. Rubino (France). Defense in 2017.

PhD in progress: Laura Aspirot, “Fluid Approximations for Stochastic Telecommunication Models”, University of the Republic, Uruguay. Advisors: E. Mordecki (Uruguay) and G. Rubino (France). Defense in 2017.

PhD in progress: Jorge Graneri, “Mathematical Models for Semantic Memory”, University of the Republic, Uruguay. Advisors: E. Mizraji (Uruguay) and G. Rubino (France). Started end 2016.

Master Project R& I 2016–2017, of Joshua Peignier and Estelle Varloot: “Game-theoretic tools to analyze classical vs collaborative economies”. Advisors: B. Tuffin and P. Maillé from Telecom Bretagne.

10.2.3. Juries

Bruno Tuffin was a member of the following PhD defense committees:

- Juan Eduardo Piccini: “Static Reliability and Resilience in Dynamic Systems”, University of the Republic, Montevideo, Uruguay, Mars 2016; reviewer.
- Maïder Estécahandy: “Méthodes accélérées de Monte-Carlo pour la simulation d’événements rares. Applications aux Réseaux de Petri”, Université de Pau et de l’Amour, Mars 2016; reviewer.
- Wenjing Shuai: “Management of electric vehicle systems with self-interested actors”, Telecom Bretagne, Sept. 2016; president.
- Simon Theissing: “Supervision en Transport Multimodal”, université de Paris-Saclay, Dec. 2016.

Bruno Sericola was a reviewer of the following PhD defense committee:

- A. Anjuka: “Stationary analysis of fluid queueing models”, Faculty of Science and Humanities, Anna University, Chennai, India.

Bruno Sericola was the president of the jury for the recruitement of Inria researchers in 2016.

Gerardo Rubino was a member of the following PhD defense committees:

- Juan Eduardo Piccini: “Static Reliability and Resilience in Dynamic Systems”, University of the Republic, Montevideo, Uruguay, Mars 2016; reviewer.
- Wassim Abbessi: “Mesures et Modélisation Fluide de Trafic Multimédia en Vue d’une Meilleure Gestion du Flux”, ENSI, Tunisia, November 12, 2016; reviewer.
- Yu-Ting Lin: “Performance analysis of video streaming services in mobile networks’, ParisTech, Paris, December 9, 2016; reviewer.

César Viho was:

- a member of juries for the recruitment of young graduate scientists and senior researchers at Inria;
- a member of the juries for the recruitment of young Associate Professors and Full Professors at ISTIC-Université Rennes 1;
- the Director of the MATISSE (Mathematics, Electronics and Computer Sciences) doctoral school in charge of managing the recruitment of PhD students and their activities during their doctorate, in all the concerned areas of the doctoral school.

10.3. Popularization

- P. Maillé and B. Tuffin responded to the public consultation on draft BEREC Guidelines on implementation of net neutrality rules. 2016.
- G. Rubino makes regular presentations to high school students about the research work in general, and specific technical topics in particular. Current talks:
  - Randomness as a tool
  - Internet as a research problem
  - Great challenges in maths: the Riemann Hypothesis
  - Great challenges in math/computer science: the “P versus NP” problem

11. Bibliography

Major publications by the team in recent years

[1] G. RUBINO, B. TUFFIN (editors). *Rare Event Simulation using Monte Carlo Methods*, John Wiley & Sons, 2009.

[2] N. BOUABDALLAH, A.-L. BEYLOT, E. DOTARO, G. PUJOLLE.*Resolving the Fairness Issues in Bus-Based Optical Access Networks*, in "IEEE Journal on Selected Areas in Communications", 2005, vol. 23, n^{o} 8, p. 1444–1457.

[3] Y. HADJADJ-AOUL, T. TALEB.*An adaptive fuzzy-based CAC scheme for uplink and downlink congestion control in converged IP and DVB-S2 networks*, in "IEEE Transactions on Wireless Communications", Feb. 2009, vol. 8, n^{o} 2, p. 816–825.

[4] Y. HAYEL, D. ROS, B. TUFFIN.*Less-than-Best-Effort Services: Pricing and Scheduling*, in "23rd IEEE Infocom Conference", Hong-Kong, China, March 2004.

[5] P. LEGUESDRON, J. PELLAUMAIL, G. RUBINO, B. SERICOLA.*Transient analysis of the M/M/1 queue*, in "Advances in Applied Probability", September 1993, vol. 25, n^{o} 3, p. 702–713.

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

[7] A. NAFAA, A. KSENTINI.*On Sustained QoS Guarantees in Operated IEEE 802.11 Wireless LANs*, in "IEEE Transactions on Parallel and Distributed Systems", 2008, vol. 19, n^{o} 8, p. 1020–1033.

[8] G. RUBINO, B. SERICOLA.*A finite characterization of weak lumpable Markov processes. Part II: The continuous time case*, in "Stochastic Processes and their Applications", 1993, vol. 45, p. 115–126.

[9] B. TUFFIN. *Bounded Normal Approximation in Highly Reliable Markovian Systems*, in "Journal of Applied Probability", 1999, vol. 36, n° 4.

Publications of the year

Articles in International Peer-Reviewed Journal

[10] E. ANCEAUME, Y. BUSNEL, E. SCHULTE-GEERS, B. SERICOLA. *Optimization Results for a Generalized Coupon Collector Problem*, in "Journal of Applied Probability", 2016, vol. 53, n° 2, https://hal.inria.fr/hal-01397403.

[11] M. BAGAA, T. TALEB, A. KSENTINI. *Efficient Tracking Area Management Framework for 5G Networks*, in "IEEE Transactions on Wireless Communications", June 2016, https://hal.inria.fr/hal-01423574.

[12] Z. I. BOTEV, P. L'ECUYER, R. SIMARD, B. TUFFIN. *Static Network Reliability Estimation under the Marshall-Olkin Copula*, in "ACM Transactions on Modeling and Computer Simulation", January 2016, vol. 26, n° 2, Article 14, https://hal.inria.fr/hal-01096393.

[13] P. FRANGOUDIS, G. C. POLYZOS, G. RUBINO. *Relay-Based Multipoint Content Delivery for Wireless Users in an Information-Centric Network*, in "Computer Networks", 2016, vol. 105, p. 207–233, https://hal.archives-ouvertes.fr/hal-01422269.

[14] R. IBRAHIM, P. L'ECUYER, H. SHEN, M. THIONGANE. *Inter-Dependent, Heterogeneous, and Time-Varying Service-Time Distributions in Call Centers*, in "European Journal of Operational Research", 2016, vol. 250, p. 480-492, https://hal.inria.fr/hal-01399541.

[15] P. L'ECUYER, C. LÉCOT, D. MUNGER, B. TUFFIN. *Sorting methods and convergence rates for Array-RQMC: some empirical comparisons*, in "Mathematics and Computers in Simulation", 2016, https://hal.inria.fr/hal-01398912.

[16] P. L'ECUYER, P. MAILLÉ, N. STIER-MOSES, B. TUFFIN. *Revenue-Maximizing Rankings for Online Platforms with Quality-Sensitive Consumers*, in "Operations Research", 2016, https://hal.inria.fr/hal-00953790.

[17] P. L'ECUYER, D. MUNGER. *LatticeBuilder: A General Software Tool for Constructing Rank-1 Lattice Rules*, in "ACM Transactions on Mathematical Software", 2016, https://hal.inria.fr/hal-01240168.

[18] P. L'ECUYER, D. MUNGER, B. ORESHKIN, R. SIMARD. *Random Numbers for Parallel Computers: Requirements and Methods*, in "Mathematics and Computers in Simulation", 2016, https://hal.inria.fr/hal-01399537.

[19] P. MAILLÉ, G. SIMON, B. TUFFIN. *Toward a Net Neutrality Debate that Conforms to the 2010s*, in "IEEE Communications Magazine", 2016, vol. 54, n° 3, p. 94-99, https://hal.inria.fr/hal-01127958.

[20] P. MAILLÉ, B. TUFFIN. *Preventing competition using side payments: when non-neutrality creates barriers to entry*, in "Netnomics", 2016 [*DOI :* 10.1007/s11066-016-9110-6], https://hal.inria.fr/hal-01398913.

[21] R. A. MARIE. *Influence of stock shortage in integrated logistic systems: an under-estimated importance*, in "International Journal of Logistics Systems and Management", June 2016, https://hal.inria.fr/hal-01419164.

[22] A. MOULOUD, D. AISSANI, Y. HADJADJ-AOUL.*Modeling and analysis data fragmentation in IEEE 802.15.4 slotted CSMA/CA protocol without ACK mode*, in "International Journal of Critical Computer-Based Systems", December 2016, https://hal.inria.fr/hal-01421547.

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

[24] A. OSAMA, A. KSENTINI.*General Model for RACH Procedure Performance Analysis*, in "IEEE Communications Letters", February 2016, https://hal.inria.fr/hal-01423570.

[25] Q. T. A. PHAM, K. PIAMRAT, K. D. SINGH, C. VIHO.*QoE-based routing algorithms for H.264/SVC video over Ad-hoc networks*, in "Wireless Networks", October 2016, vol. 22, n^{o} 7 [*DOI :* 10.1007/s11276-015-1103-0], https://hal.inria.fr/hal-01253420.

[26] P. T. A. QUANG, K. PIAMRAT, K. D. SINGH, C. VIHO.*Video Streaming over Ad-hoc Networks: a QoE-based Optimal Routing Solution*, in "IEEE Transactions on Vehicular Technology", April 2016 [*DOI :* 10.1109/TVT.2016.2552041], https://hal.archives-ouvertes.fr/hal-01340989.

[27] G. RUBINO, H. CANCELA, L. MURRAY.*Conditional Monte Carlo with Intermediate Estimations for simulation of Markovian systems*, in "Electronic Notes in Theoretical Computer Science", 2016, vol. 321, p. 3 - 21, https://hal.inria.fr/hal-01261168.

[28] G. RUBINO, F. R. PADULA, P. ROMERO, A. MAUTTONE.*Capacitated m Ring Star Problem under Diameter Constrained Reliability*, in "Electronic Notes in Theoretical Computer Science", 2016, https://hal.inria.fr/hal-01261161.

[29] T. TALEB, A. KSENTINI, M. CHEN, R. JANTTI.*Coping with Emerging Mobile Social Media Applications through Dynamic Service Function Chaining*, in "IEEE Transactions on Wireless Communications", April 2016, https://hal.inria.fr/hal-01423572.

[30] T. TALEB, A. KSENTINI, P. FRANGOUDIS.*Follow-Me Cloud: When Cloud Services Follow Mobile Users*, in "IEEE Transactions on Cloud Computing", 2016, https://hal.archives-ouvertes.fr/hal-01422271.

[31] T. TALEB, A. KSENTINI, B. SERICOLA.*On Service Resilience in Cloud-Native 5G Mobile Systems*, in "IEEE Journal on Selected Areas in Communications", 2016, vol. 34, n^{o} 3, p. 483 - 496 [*DOI :* 10.1109/JSAC.2016.2525342], https://hal.inria.fr/hal-01398157.

Invited Conferences

[32] M. BOUZOUITA, Y. HADJADJ-AOUL, N. ZANGAR, S. TABBANE.*Leveraging crowd sourcing technique for an optimized M2M access during emergency situations*, in "International Conference on Information and Communication Technologies for Disaster Management (ICT-DM 2016)", Vienna, Austria, IEEE (editor), IEEE, December 2016, https://hal.inria.fr/hal-01421787.

[33] P. MAILLÉ, B. TUFFIN.*Auctions For Online Ad Space Among Advertisers Mixing Pay-per-view And Pay-per-click*, in "Informs Annual Meeting", Nashville, United States, November 2016, https://hal.inria.fr/hal-01398934.

[34] G. RUBINO.*Effectiveness: a new metric in performance evaluation: Keynote presentation*, in "Modeling, Analysis and Simulation of Computer and Telecommunication Systems", Londres, United Kingdom, Proceedings of Mascots'16, IEEE, September 2016, Keynote, https://hal.inria.fr/hal-01423833.

[35] G. RUBINO.*Mesure automatique de la qualité perçue et applications : La technologie PSQA de l'Inria*, in "Technoconférence "Nouvelles expériences son et vidéo"", Rennes, France, Pôle de compétitivité "Images et Réseaux", October 2016, https://hal.inria.fr/hal-01424095.

[36] G. RUBINO.*Perceptual Quality estimation: Machine Learning and Big Data problems: Keynote*, in "8th International Conference on Soft Computing and Pattern Recognition (SoCPaR 2016)", Vellore, India, SoCPaR'16, VIT, Vellore, December 2016, https://hal.inria.fr/hal-01424098.

[37] G. RUBINO.*Static models, recursive estimators and the zero-variance approach: Invited talk*, in "Advances in Uncertainty Quantification Methods, algorithms and Applications", Thuwal, Saudi Arabia, UQ Center, Kaust, Saudi-Arabia, January 2016, https://hal.inria.fr/hal-01424093.

[38] G. RUBINO.*Two Big Data projects in telecommunications: Keynote presentation*, in "Math AmSud Seminar", Montevideo, Uruguay, ANII, Uruguay, November 2016, https://hal.inria.fr/hal-01424096.

[39] G. RUBINO.*Zero-variance Approaches in Static Reliability Problems: Invited talk*, in "SIAM Conference on Uncertainty Quantification", Lausanne, Switzerland, SIAM Conference on Uncertainty Quantification, SIAM, April 2016, https://hal.inria.fr/hal-01424094.

[40] B. TUFFIN.*Some applications of importance sampling to dependability analysis*, in "International Conference on Monte Carlo techniques", Paris, France, July 2016, https://hal.inria.fr/hal-01398928.

International Conferences with Proceedings

[41] E. ANCEAUME, Y. BUSNEL, N. RIVETTI, B. SERICOLA.*Identifier des icebergs parmi des flux de données distribués*, in "ALGOTEL 2016 - 18èmes Rencontres Francophones sur les Aspects Algorithmiques des Télécommunications", Bayonne, France, ALGOTEL 2016 - 18èmes Rencontres Francophones sur les Aspects Algorithmiques des Télécommunications, May 2016, https://hal.archives-ouvertes.fr/hal-01303873.

[42] E. ANCEAUME, T. LAJOIE-MAZENC, R. LUDINARD, B. SERICOLA.*Safety Analysis of Bitcoin Improvement Proposals*, in "IEEE Symposium on Network Computing and Applications", Boston, United States, IEEE, October 2016, https://hal.archives-ouvertes.fr/hal-01397685.

[43] E. ANCEAUME, R. LUDINARD, B. SERICOLA.*L'empire romain ne doit pas être géré comme une petite île grecque*, in "ALGOTEL 2016 - 18èmes Rencontres Francophones sur les Aspects Algorithmiques des Télécommunications", Bayonne, France, May 2016, https://hal.archives-ouvertes.fr/hal-01305334.

[44] M. BAGAA, A. KSENTINI, T. TALEB, R. JANTTI, A. CHELLI, I. BALASINGHAM.*An efficient D2D-based strategies for Machine Type Communications in 5G mobile systems*, in "IEEE Wireless Communications & Networking Conference 2016", Doha, Qatar, April 2016, https://hal.inria.fr/hal-01423579.

[45] H. BEN AMMAR, Y. HADJADJ-AOUL, S. AÏT-CHELLOUCHE.*CLIC: a CLIque-based cooperative Caching for Content Centric Networking*, in "Global Information Infrastructure and Networking Symposium (GIIS 2016)", Porto, Portugal, IEEE (editor), October 2016, https://hal.inria.fr/hal-01421595.

[46] Z. I. BOTEV, P. L'ECUYER.*Simulation from the Normal Distribution Truncated to an Interval in the Tail*, in "ValueTools 2016", Taormina, Italy, October 2016, https://hal.inria.fr/hal-01399506.

[47] Z. I. BOTEV, P. L'ECUYER, B. TUFFIN.*Permutation Monte Carlo : applications, analysis and extension*, in "11th International Workshop on Rare Event Simulation (RESIM 2016)", Eindhoven, Netherlands, April 2016, https://hal.inria.fr/hal-01398933.

[48] M. BOUZOUITA, Y. HADJADJ-AOUL, N. ZANGAR, G. RUBINO, S. TABBANE.*Applying nonlinear optimal control strategy for the access management of MTC devices*, in "13th Annual IEEE Consumer Communications & Networking (CCNC)", Las Vegas, United States, January 2016, https://hal.inria.fr/hal-01252011.

[49] *Best Paper*
M. BOUZOUITA, Y. HADJADJ-AOUL, N. ZANGAR, G. RUBINO, S. TABBANE.*Dynamic adaptive access barring scheme for heavily congested M2M networks*, in "International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems (MSWIM 2016)", Malta, Malta, ACM (editor), ACM, November 2016, https://hal.inria.fr/hal-01421614.

[50] M. BOUZOUITA, Y. HADJADJ-AOUL, N. ZANGAR, S. TABBANE.*Adaptive access protocol for heavily congested M2M networks*, in "IEEE Symposium on Computers and Communications (ISCC 2016)", Messina, Italy, IEEE (editor), IEEE, June 2016, https://hal.inria.fr/hal-01421620.

[51] M. BOUZOUITA, Y. HADJADJ-AOUL, N. ZANGAR, S. TABBANE.*On the risk of congestion collapse in heavily congested M2M networks*, in "International Symposium on Networks, Computers and Communications (ISNCC 2016)", Hammamet, Tunisia, May 2016, https://hal.inria.fr/hal-01421793.

[52] E. BTISSAM, A. KSENTINI, J.-E. MEDDOUR.*A Traffic-Driven Analysis for Small Cells Backhaul Planning*, in "IEEE Wireless Communications & Networking Conference 2016", Doha, Qatar, April 2016, https://hal.inria.fr/hal-01423582.

[53] E. CANALE, P. ROMERO, G. RUBINO.*Factorization theory in diameter constrained reliability*, in "2016 8th International Workshop on Resilient Networks Design and Modeling (RNDM)", Halmstad, Sweden, September 2016, https://hal.inria.fr/hal-01423583.

[54] E. CANALE, P. ROMERO, G. RUBINO, W. XAVIER.*Network utility problem and easy reliability polynomials*, in "2016 8th International Workshop on Resilient Networks Design and Modeling (RNDM)", Halmstad, Sweden, September 2016, https://hal.inria.fr/hal-01423584.

[55] H. CANCELA, L. MURRAY, G. RUBINO.*Highly Reliable Stochastic Flow Network Reliability Estimation*, in "CLEI'16 (Latino-American Computer Science conference XLII)", Valparaiso, Chile, October 2016, https://hal.inria.fr/hal-01423588.

[56] W. CHANG, T. A. TA, P. L'ECUYER, F. BASTIN.*Two-stage chance-constrainted staffing with agent recourse for multi-skill call centers*, in "Winter Simulation Conference", Arlington, United States, December 2016, https://hal.inria.fr/hal-01399534.

[57] S. DUTTA, T. TALEB, P. FRANGOUDIS, A. KSENTINI.*On-the-Fly QoE-Aware Transcoding in the Mobile Edge*, in "Proc. IEEE Globecom", 2016, https://hal.archives-ouvertes.fr/hal-01422268.

[58] S. DUTTA, T. TALEB, A. KSENTINI.*QoE-aware elasticity support in cloud-native 5G systems*, in "IEEE International Conference on Communications (ICC 2016)", Kuala Lampur, Malaysia, May 2016, https://hal.inria.fr/hal-01423575.

[59] P. FRANGOUDIS, L. YALA, A. KSENTINI, T. TALEB.*An architecture for on-demand service deployment over a telco CDN*, in "Proc. IEEE ICC", 2016, https://hal.archives-ouvertes.fr/hal-01422266.

[60] F. GRANDHOMME, G. GUETTE, A. KSENTINI, T. PLESSE.*Comparing inter-domain routing protocol assessment tools for MANET*, in "2016 IEEE International Conference on Communications (ICC)", Kuala Lumpur, Malaysia, 2016 IEEE International Conference on Communications (ICC), May 2016, https://hal.inria.fr/hal-01355402.

[61] F. GRANDHOMME, G. GUETTE, A. KSENTINI, T. PLESSE.*ITMAN: An Inter Tactical Mobile Ad Hoc Network Routing Protocol*, in "MILCOM2016", Baltimore, United States, November 2016, https://hal.inria.fr/hal-01397710.

[62] D. E. HENNI, Y. HADJADJ-AOUL, A. GHOMARI.*Probe-SDN: a smart monitoring framework for SDN-based networks*, in "Global Information Infrastructure and Networking Symposium (GIIS 2016)", Porto, Portugal, IEEE (editor), IEEE, October 2016, https://hal.inria.fr/hal-01421599.

[63] N. HERBAUT, D. NEGRU, D. MAGONI, P. A. FRANGOUDIS.*Deploying a content delivery service function chain on an SDN-NFV operator infrastructure*, in "6th International Conference on Telecommunications and Multimedia (TEMU)", Heraklion, Greece, Proceedings of the 6th International Conference on Telecommunications and Multimedia (TEMU), July 2016 [*DOI :* 10.1109/TEMU.2016.7551917], https://hal.archives-ouvertes.fr/hal-01422265.

[64] N. HERBAUT, D. NÉGRU, Y. CHEN, P. FRANGOUDIS, A. KSENTINI.*Content Delivery Networks as a Virtual Network Function: a Win-win ISP-CDN Collaboration*, in "Proc. IEEE Globecom", 2016, https://hal.archives-ouvertes.fr/hal-01422267.

[65] N. JARA, G. RUBINO, R. VALLEJOS.*Blocking evaluation of dynamic WDM networks without wavelength conversion*, in "2016 21st European Conference on Networks and Optical Communications (NOC)", Lisbon, Portugal, June 2016, https://hal.inria.fr/hal-01423585.

[66] A. KSENTINI, M. BAGAA, T. TALEB, I. BALASINGHAM.*On using bargaining game for Optimal Placement of SDN controllers*, in "IEEE International Conference on Communications (ICC 2016)", Kuala Lampur, Malaysia, May 2016, https://hal.inria.fr/hal-01423576.

[67] P. L'ECUYER, P. MAILLÉ, N. STIER-MOSES, B. TUFFIN.*Search (Non-)Neutrality and Impact on Innovation*, in "NetEcon 2016 : The 11th Workshop on the Economics of Networks, Systems and Computation", Juan-les-Pins, France, June 2016, https://hal.inria.fr/hal-01330741.

[68] R. LEONE, F. SISMONDI, T. WATTEYNE, C. VIHO.*Technical Overview of F-Interop*, in "Conference on Interoperability in IoT (InterIoT)", Paris, France, October 2016, https://hal.inria.fr/hal-01363391.

[69] Y. LI, P. FRANGOUDIS, Y. HADJADJ-AOUL, P. BERTIN.*A Mobile-Edge-Computing-Based Architecture for Improved Adaptive HTTP Video Delivery*, in "IEEE Conference on Standards for Communications and Networking (CSCN)", Berlin, Germany, IEEE, October 2016, https://hal.inria.fr/hal-01421583.

[70] P. MAILLÉ, G. SIMON, B. TUFFIN.*On Revenue-oriented Content Delivery Networks And Their Impact On Net Neutrality*, in "Informs Annual Meeting", Nashville, United States, November 2016, https://hal.inria.fr/hal-01398937.

[71] P. MAILLÉ, G. SIMON, B. TUFFIN.*Vertical Integration of CDN and Network Operator: Model and Analysis*, in "IEEE/ACM International Symposium on Modelling, Analysis and Simulation of Computer and Telecommunication Systems (MASCOTS 2016)", London, United Kingdom, September 2016, https://hal.inria.fr/hal-01398917.

[72] R. A. MARIE.*Should I Add an Intermittent Server*, in "10th EAI International Conference on Performance Evaluation Methodologies and Tools (VALUETOOLS'16)", Taormina, Italy, July 2016, https://hal.inria.fr/hal-01419167.

[73] O. MATUS, J. BARRERA, E. MORENO, G. RUBINO.*Calibrating a Dependent Failure Model for Computing Reliabilities in Telecommunication Networks*, in "Winter Simulation Conference", Arlington, United States, December 2016, https://hal.inria.fr/hal-01423587.

[74] Y. MOCQUARD, E. ANCEAUME, B. SERICOLA.*Optimal Proportion Computation with Population Protocols*, in "Symposium on Network Computing and Applications", Boston, United States, IEEE, October 2016, https://hal.archives-ouvertes.fr/hal-01354352.

[75] *Best Paper*
Y. MOCQUARD, B. SERICOLA, S. ROBERT, E. ANCEAUME.*Analysis of the Propagation Time of a Rumour in Large-scale Distributed Systems*, in "Symposium on Network Computing and Applications", Boston, United States, October 2016, This article has received the Best Student Paper Award, https://hal.archives-ouvertes.fr/hal-01354815.

[76] A. RAI, R. C. VALENZUELA, B. TUFFIN, G. RUBINO, P. DERSIN.*Approximate zero-variance importance sampling for static network reliability estimation with node failures and application to rail systems*, in "Winter Simulation Conference", Arlington, United States, December 2016, https://hal.inria.fr/hal-01398921.

[77] N. RIVETTI, E. ANCEAUME, Y. BUSNEL, L. QUERZONI, B. SERICOLA.*Online Scheduling for Shuffle Grouping in Distributed Stream Processing Systems Research Paper*, in "ACM/IFIP/USENIX Middleware 2016", Trento, Italy, ACM/IFIP/USENIX, December 2016 [*DOI :* 10.1145/2988336.2988347], https://hal.archives-ouvertes.fr/hal-01397658.

[78] N. RIVETTI, Y. BUSNEL, A. MOSTEFAOUI.*Résumer efficacement des flux de données massifs en fenêtre glissante*, in "ALGOTEL 2016 - 18èmes Rencontres Francophones sur les Aspects Algorithmiques des Télécommunications", Bayonne, France, ALGOTEL 2016 - 18èmes Rencontres Francophones sur les Aspects Algorithmiques des Télécommunications, May 2016, https://hal.archives-ouvertes.fr/hal-01303882.

[79] N. RIVETTI, Y. BUSNEL, L. QUERZONI.*Load-aware shedding in stream processing systems*, in "10th ACM International Conference on Distributed and Event-based Systems (DEBS)", Irvine, United States, June 2016, p. 61 - 68 [*DOI :* 10.1145/2933267.2933311], https://hal.inria.fr/hal-01413212.

[80] N. RIVETTI, L. QUERZONI, E. ANCEAUME, Y. BUSNEL, B. SERICOLA.*Groupement de clés efficace pour un équilibrage de charge quasi-optimal dans les systèmes de traitement de flux*, in "ALGOTEL 2016 - 18èmes

Rencontres Francophones sur les Aspects Algorithmiques des Télécommunications", Bayonne, France, ALGOTEL 2016 - 18èmes Rencontres Francophones sur les Aspects Algorithmiques des Télécommunications, May 2016, https://hal.archives-ouvertes.fr/hal-01303887.

[81] A. SAMBA, Y. BUSNEL, A. BLANC, P. DOOZE, G. SIMON.*Throughput Prediction in Cellular Networks: Experiments and Preliminary Results*, in "1ères Rencontres Francophones sur la Conception de Protocoles, l'Évaluation de Performance et l'Expérimentation des Réseaux de Communication (CoRes 2016)", Bayonne, France, 1ères Rencontres Francophones sur la Conception de Protocoles, l'Évaluation de Performance et l'Expérimentation des Réseaux de Communication, May 2016, https://hal.archives-ouvertes.fr/hal-01311158.

[82] J.-M. SANNER, Y. HADJADJ-AOUL, M. OUZZIF, G. RUBINO.*Hierarchical clustering for an efficient controllers' placement in software defined networks*, in "Global Information Infrastructure and Networking Symposium (GIIS 2016)", Porto, Portugal, IEEE (editor), IEEE, October 2016, https://hal.inria.fr/hal-01421587.

[83] S. SHANMUGALINGAM, A. KSENTINI, P. BERTIN.*DPDK Open vSwitch performance validation with mirroring feature*, in "International Conference on Telecommunications (ICT)", Thessaloniki, Greece, May 2016, https://hal.inria.fr/hal-01423578.

[84] T. A. TA, P. L'ECUYER, F. BASTIN.*Staffing optimization with chance constraints for emergency call centers*, in "MOSIM 2016 - 11th International Conference on Modeling, Optimization and Simulation", Montréal, Canada, August 2016, https://hal.inria.fr/hal-01399507.

[85] T. TALEB, K. SAMDANIS, A. KSENTINI.*Towards Elastic Application-oriented Bearer Management for enhancing QoE in LTE Networks*, in "IEEE Wireless Communications & Networking Conference 2016", Doha, Qatar, April 2016, https://hal.inria.fr/hal-01423580.

[86] M. THIONGANE, W. CHANG, P. L'ECUYER.*New History-Based Delay Predictors for Service Systems*, in "Winter Simulation Conference", Arlington, United States, December 2016, https://hal.inria.fr/hal-01399535.

[87] L. YALA, P. FRANGOUDIS, A. KSENTINI.*QoE-Aware Computing Resource Allocation for CDN-as-a-Service Provision*, in "Proc. IEEE Globecom", 2016, https://hal.archives-ouvertes.fr/hal-01422270.

[88] S. ZIEGLER, S. FDIDA, T. WATTEYNE, C. VIHO.*F-Interop - Online Conformance, Interoperability and Performance Tests for the IoT*, in "Conference on Interoperability in IoT (InterIoT)", Paris, France, October 2016, https://hal.inria.fr/hal-01363427.

Conferences without Proceedings

[89] E. ANCEAUME, R. LUDINARD, B. SERICOLA.*Relying on Consensus does not Make Bitcoin Safer*, in "Fast Abstract in the 46th Annual IEEE/IFIP International Conference on Dependable Systems and Networks", Toulouse, France, M. ROY, J. A. LOPEZ, A. CASIMIRO (editors), DSN2016-FAST-ABSTRACT, June 2016, https://hal.archives-ouvertes.fr/hal-01316541.

[90] F. GRANDHOMME, G. GUETTE, A. KSENTINI, T. PLESSE.*Comparaison d'outils d'évaluation de performance des protocoles de routage inter-MANET*, in "13ème Conference Francophone sur les Nouvelles Technologies de la Répartition (NOTERE 2016)", Paris, France, July 2016, https://hal.inria.fr/hal-01355407.

[91] Y. LI, Y. HADJADJ-AOUL, P. BERTIN, G. RUBINO.*A control theoretic strategy for intelligent network interface selection in heterogeneous environments*, in "13th Annual IEEE Consumer Communications & Networking", Las Vegas, United States, IEEE, January 2016, https://hal.inria.fr/hal-01252008.

[92] P. MAILLÉ, B. TUFFIN.*Enchères pour espaces publicitaires avec annonceurs sensibles aux vues et aux clics*, in "17ème congrès de la Société Française de Recherche Opérationnelle et d'Aide A la Decision (ROADEF)", Compiègne, France, February 2016, https://hal.inria.fr/hal-01398939.

[93] G. MOUALLA, P. FRANGOUDIS, Y. HADJADJ-AOUL, S. AÏT-CHELLOUCHE.*A bloom-filter-based socially aware scheme for content replication in mobile ad hoc networks*, in "13th Annual IEEE Consumer Communications & Networking", Las Vegas, United States, IEEE, January 2016 [*DOI :* 10.1109/CCNC.2016.7444807], https://hal.inria.fr/hal-01252001.

[94] Q. T. A. PHAM, K. PIAMRAT, K. D. SINGH, C. VIHO.*Q-RoSA: QoE-aware routing for SVC video streaming over ad-hoc networks*, in "13th IEEE Annual Consumer Communications & Networking Conference (CCNC)", Las Vegas, United States, January 2016 [*DOI :* 10.1109/CCNC.2016.7444863], https://hal.archives-ouvertes.fr/hal-01341209.

[95] Q. T. A. PHAM, K. PIAMRAT, K. D. SINGH, C. VIHO.*Q-SWiM: QoE-based Routing algorithm for SVC Video Streaming over Wireless Mesh Networks*, in "PIMRC 2016 - IEEE 27th Annual International Symposium on Personal and Indoor Mobile Radio Conference", Valencia, Spain, September 2016, https://hal.archives-ouvertes.fr/hal-01386166.

Scientific Books (or Scientific Book chapters)

[96] R. A. MARIE.*Introduction aux probabilités - Applications en informatique et télécommunications, fiabilité et gestion d'incertitudes*, Ellipses, June 2016, https://hal.inria.fr/hal-01419165.

[97] R. A. MARIE.*From Performability to Uncertainty*, in "Principles of Performance and Reliability Modeling and Evaluation", L. FIONDELLA, A. PULIAFITO (editors), Springer International, June 2016, p. 143-168, https://hal.inria.fr/hal-01419166.

[98] G. RUBINO, B. SERICOLA.*Sojourn Times in Dependability Modeling*, in "Principles of Performance and Reliability Modeling and Evaluation", L. FIONDELLA, A. PULIAFITO (editors), Springer International, 2016, https://hal.inria.fr/hal-01258121.

Books or Proceedings Editing

[99] Y. HADJADJ-AOUL, S. LAHOUD (editors). *Information and Communication Technologies for Disaster Management*, International Journal of Distributed Systems and Technologies (IJDST), IGI Global, October 2016, vol. 7, n^{o} 4, https://hal.inria.fr/hal-01421565.

[100] J.-H. LEE, K. D. SINGH, Y. HADJADJ-AOUL, N. KUMAR (editors). *Wireless and Mobile Technologies for the Internet of Things*, Mobile Information Systems, Hindawi, February 2016, https://hal.inria.fr/hal-01421354.

Research Reports

[101] E. ANCEAUME, F. CASTELLA, A. MOSTEFAOUI, B. SERICOLA.*Performance Evaluation of a Distributed Synchronization Protocol*, Inria ; Irisa ; Lina ; Irmar, March 2016, https://hal.inria.fr/hal-01283064.

[102] N. RIVETTI, E. ANCEAUME, Y. BUSNEL, L. QUERZONI, B. SERICOLA.*Load-Aware Shedding in Stream Processing Systems*, LINA-University of Nantes ; Sapienza Università di Roma (Italie) ; Irisa ; Inria Rennes, May 2016, https://hal.inria.fr/hal-01311970.

Patents and standards

[103] Y. LI, Y. HADJADJ-AOUL, P. BERTIN.*Optimisation des flux DASH dans les réseaux mobiles 4G et au-delà*, April 2016, n^{o} 0, https://hal.inria.fr/hal-01421365.

Other Publications

[104] E. GOURDIN, P. MAILLÉ, G. SIMON, B. TUFFIN.*The Economics of CDNs and Their Impact on Neutrality*, November 2016, working paper or preprint, https://hal.inria.fr/hal-01398923.

[105] P. MAILLÉ, B. TUFFIN.*Non-Neutrality Pushed by Big Content Providers*, November 2016, working paper or preprint, https://hal.inria.fr/hal-01398926.

[106] P. MAILLÉ, B. TUFFIN.*Enforcing free roaming among UE countries: an economic analysis*, 2017, working paper or preprint, https://hal.inria.fr/hal-01428920.

[107] J.-M. SANNER, M. OUZZIF, Y. HADJADJ-AOUL, J.-E. DARTOIS.*Evolutionary algorithms for optimized SDN controllers & NVFs' placement in SDN networks*, November 2016, SDN Day 2016, Poster, https://hal.inria.fr/hal-01421799.

References in notes

[108] G. RUBINO, B. TUFFIN (editors). *Rare Event Simulation using Monte Carlo Methods*, John Wiley & Sons, 2009.

[109] P. MAILLÉ, B. TUFFIN.*Telecommunication Network Economics: From Theory to Applications*, Cambridge University Press, 2014.

[110] G. RUBINO.*Power and effectiveness in queueing systems*, in "12th International Conference on Quantitative Evaluation of SysTems (QEST'15)", Universidad Complutense de Madrid, Spain, IEEE CS Press, September 2015.

[111] G. RUBINO, B. SERICOLA.*Markov Chains and Dependability Theory*, Cambridge University Press, 2014.

Project-Team DIVERSE

Diversity-centric Software Engineering

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

IN PARTNERSHIP WITH:
Institut national des sciences appliquées de Rennes

Université Rennes 1

RESEARCH CENTER
Rennes - Bretagne-Atlantique

THEME
Distributed programming and Software engineering

Table of contents

Project-Team DIVERSE

Creation of the Team: 2014 January 01, updated into Project-Team: 2014 July 01

Keywords:

Computer Science and Digital Science:
1.2.1. - Dynamic reconfiguration
2.1.2. - Object-oriented programming
2.1.10. - Domain-specific languages
2.5. - Software engineering
2.5.2. - Component-based Design
2.5.3. - Empirical Software Engineering
2.6.2. - Middleware
4.4. - Security of equipment and software
4.8. - Privacy-enhancing technologies

Other Research Topics and Application Domains:
3.1. - Sustainable development
3.1.1. - Resource management
6.1. - Software industry
6.1.1. - Software engineering
6.4. - Internet of things
6.5. - Information systems
6.6. - Embedded systems
8.1.2. - Sensor networks for smart buildings
9.4.1. - Computer science

1. Members

Research Scientist
Benoit Baudry [Team leader, Inria, Researcher, HDR]

Faculty Members
Mathieu Acher [Univ. Rennes I, Associate Professor]
Olivier Barais [Univ. Rennes I, Professor, HDR]
Arnaud Blouin [INSA Rennes, Associate Professor]
Johann Bourcier [Univ. Rennes I, Associate Professor]
Benoit Combemale [Univ. Rennes I, Associate Professor, HDR]
Jean-Marc Jezequel [Univ. Rennes I, Professor, HDR]
Noel Plouzeau [Univ. Rennes I, Associate Professor]

Technical Staff
Simon Allier [Inria, until Nov 2016]
Fabien Coulon [Univ. Rennes I]
Andre Elie [Univ. Rennes I]
Jose Angel Galindo Duarte [Inria, until Sep 2016]
Caroline Landry [Inria, from Dec 2016]
Manuel Leduc [Inria]

Dorian Leroy [Univ. Rennes I]
Jean Parpaillon [Inria, until Aug 2016]
Maxime Tricoire [Inria]

PhD Students

Sana Ben Nasr [Inria, until Apr 2016]
Mohamed Boussaa [Univ. Rennes I]
Kevin Corre [Orange Labs, granted by CIFRE]
Thomas Degueule [Univ. Rennes I, until Nov 2016]
Alejandro Gomez Boix [Inria, from Nov 2016]
Nicolas Harrand [Inria, from Nov 2016]
Pierre Laperdrix [Univ. Rennes I]
Gwendal Le Moulec [INSA Rennes]
David Mendez Acuna [Inria, until Oct 2016]
Johan Pelay [Inst. de Recherche Technologique B-COM, until Mar 2016]
Quentin Plazar [Inria]
Oscar Luis Vera Perez [Inria]
Alexandre Rio [OKWind]
Jean-Emile Dartois [B-COM]
Youssou Ndaye [Orange]
Marcelino Rodriguez Cancio [Univ. Rennes I]
Paul Temple [Univ. Rennes I]
Kwaku Yeboah-Antwi [Inria, until Sep 2016]
Guillaume Becan [Univ. Rennes I, intern]

Visiting Scientists

Marwa Hentati [National School of Engineers of Sfax (ENIS), from Apr 2016 until May 2016]
Akram Kammoun ["Research laboratory on Development and Control of Distributed Applications (ReD-CAD)" Sfax Tunisia, from May 2016 until Jun 2016]

Administrative Assistant

Tifenn Donguy [CNRS]

Others

Nassim Aliche [Univ. Rennes I, intern, from Jul 2016 until Aug 2016]
Marvin Billaud [Univ. Rennes I, intern, from May 2016 until Aug 2016]
Clementine Delambily [Inria, intern, from May 2016 until Aug 2016]
Pierre Duros [Univ. Rennes I, intern, from Jul 2016 until Sep 2016]
Laureline Guerin [Inria, intern, until Aug 2016]
Axel Halin [Univ. Rennes I, intern, from Sep 2016]
Remi Hauperpin Bonneau [Univ. Rennes I, intern, from May 2016 until Aug 2016]
Pierre Le Gall [Univ. Rennes I, intern, from Feb 2016 until Jun 2016]
Gurvan Le Guernic [DGA, research engineer]
Stephane Mangin [Univ. Rennes I, intern, until Aug 2016]
Malo Manini [Univ. Rennes I, intern, from Jun 2016 until Aug 2016]
Bruno Merciol [Univ. Rennes I, intern, until Jun 2016]
Vikas Mishra [Inria, intern, from May 2016 until Jul 2016]
Leo Noel-Baron [Univ. Rennes I, intern, from May 2016 until Jul 2016]
Alexandre Nuttinck [Univ. Rennes I, intern, from Sep 2016]

2. Overall Objectives

2.1. Overall objectives

DIVERSE's research agenda is in the area of software engineering. In this broad domain we develop models, methodologies and theories to address the challenges raised by the emergence of several forms of diversity in the design, deployment and evolution of software-intensive systems. The emergence of software diversity is an essential phenomenon in all application domains that we investigate with our industrial partners. These application domains range from complex systems such as systems of systems (in collaboration with Thales and DGA) and Instrumentation and Control (with EDF) to pervasive combinations of Internet of Things and Internet of Services (with TellU and Software AG) and tactical information systems (with the firefighter department). Even if today these systems are apparently radically different, we envision a strong convergence of the scientific principles underpinning their construction and validation towards **flexible and open yet dependable systems**. In particular, we see that the required flexibility and openness raise challenges for the software layer of these systems that must deal with four dimensions of diversity: the **diversity of languages** used by the stakeholders involved in the construction of these systems; the **diversity of features** required by the different customers; the **diversity of runtime environments** in which software has to run and adapt; the **diversity of implementations** that are necessary for resilience through redundancy.

In this context, the major software engineering challenge consists in handling **diversity** from variability in requirements and design to heterogeneous and dynamic execution environments. In particular this requires considering that the software system must adapt, in unpredictable ways, to changes in the requirements and environment. Conversely, explicitly handling of diversity is a great opportunity to allow software to spontaneously explore alternative design solutions. Concretely, we want to provide software engineers with the ability:

- to characterize an 'envelope' of possible variations
- to compose 'envelopes' (to discover new macro envelopes in an opportunistic manner)
- to dynamically synthesize software inside a given envelop

The major scientific objective that we must achieve to provide such mechanisms for software engineering is synthesized below

Scientific objective for DIVERSE: Automatically **compose and synthesize software diversity** from design to runtime to **address unpredictable evolutions of software-intensive systems**

Software product lines and associated variability modeling formalisms represent an essential aspect of software diversity, which we already explored in the past and that represent a major foundation of DIVERSE's research agenda. However, DIVERSE also exploits other foundations to handle new forms of diversity: type theory and models of computation for the composition of languages; distributed algorithms and pervasive computation to handle the diversity of execution platforms; functional and qualitative randomized transformations to synthesize diversity for robust systems.

3. Research Program

3.1. Scientific background

3.1.1. Model-driven engineering

Model-Driven Engineering (MDE) aims at reducing the accidental complexity associated with developing complex software-intensive systems (e.g., use of abstractions of the problem space rather than abstractions of the solution space) [117]. It provides DIVERSE with solid foundations to specify, analyze and reason about the different forms of diversity that occur through the development lifecycle. A primary source of accidental complexity is the wide gap between the concepts used by domain experts and the low-level abstractions provided by general-purpose programming languages [88]. MDE approaches address this problem through modeling techniques that support separation of concerns and automated generation of major system artifacts from models (*e.g.,* test cases, implementations, deployment and configuration scripts). In MDE, a model describes an aspect of a system and is typically created or derived for specific development purposes [70]. Separation of concerns is supported through the use of different modeling languages, each providing constructs based on abstractions that are specific to an aspect of a system. MDE technologies also provide support for manipulating models, for example, support for querying, slicing, transforming, merging, and analyzing (including executing) models. Modeling languages are thus at the core of MDE, which participates to the development of a sound *Software Language Engineering*[0], including an unified typing theory that integrate models as first class entities [120].

Incorporating domain-specific concepts and high-quality development experience into MDE technologies can significantly improve developer productivity and system quality. Since the late nineties, this realization has led to work on MDE language workbenches that support the development of domain-specific modeling languages (DSMLs) and associated tools (*e.g.,* model editors and code generators). A DSML provides a bridge between the field in which domain experts work and the implementation (programming) field. Domains in which DSMLs have been developed and used include, among others, automotive, avionics, and the emerging cyber-physical systems. A study performed by Hutchinson et al. [94] provides some indications that DSMLs can pave the way for wider industrial adoption of MDE.

More recently, the emergence of new classes of systems that are complex and operate in heterogeneous and rapidly changing environments raises new challenges for the software engineering community. These systems must be adaptable, flexible, reconfigurable and, increasingly, self-managing. Such characteristics make systems more prone to failure when running and thus the development and study of appropriate mechanisms for continuous design and run-time validation and monitoring are needed. In the MDE community, research is focused primarily on using models at design, implementation, and deployment stages of development. This work has been highly productive, with several techniques now entering a commercialization phase. As software systems are becoming more and more dynamic, the use of model-driven techniques for validating and monitoring run-time behavior is extremely promising [102].

3.1.2. Variability modeling

While the basic vision underlying *Software Product Lines* (SPL) can probably be traced back to David Parnas seminal article [110] on the Design and Development of Program Families, it is only quite recently that SPLs are emerging as a paradigm shift towards modeling and developing software system families rather than individual systems [108]. SPL engineering embraces the ideas of mass customization and software reuse. It focuses on the means of efficiently producing and maintaining multiple related software products, exploiting what they have in common and managing what varies among them.

Several definitions of the *software product line* concept can be found in the research literature. Clements *et al.* define it as *a set of software-intensive systems sharing a common, managed set of features that satisfy the specific needs of a particular market segment or mission and are developed from a common set of core assets in a prescribed way* [107]. Bosch provides a different definition [76]: *A SPL consists of a product line architecture and a set of reusable components designed for incorporation into the product line architecture. In addition, the PL consists of the software products developed using the mentioned reusable assets*. In spite of the similarities, these definitions provide different perspectives of the concept: *market-driven*, as seen by Clements *et al.*, and *technology-oriented* for Bosch.

SPL engineering is a process focusing on capturing the *commonalities* (assumptions true for each family member) and *variability* (assumptions about how individual family members differ) between several software products [82]. Instead of describing a single software system, a SPL model describes a set of products in the same domain. This is accomplished by distinguishing between elements common to all SPL members, and those that may vary from one product to another. Reuse of core assets, which form the basis of the product line, is key to productivity and quality gains. These core assets extend beyond simple code reuse and may

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

include the architecture, software components, domain models, requirements statements, documentation, test plans or test cases.

The SPL engineering process consists of two major steps:

1. **Domain Engineering**, or *development for reuse*, focuses on core assets development.
2. **Application Engineering**, or *development with reuse*, addresses the development of the final products using core assets and following customer requirements.

Central to both processes is the management of **variability** across the product line [90]. In common language use, the term *variability* refers to *the ability or the tendency to change*. Variability management is thus seen as the key feature that distinguishes SPL engineering from other software development approaches [77]. Variability management is thus growingly seen as the cornerstone of SPL development, covering the entire development life cycle, from requirements elicitation [122] to product derivation [127] to product testing [106], [105].

Halmans *et al.* [90] distinguish between *essential* and *technical* variability, especially at requirements level. Essential variability corresponds to the customer's viewpoint, defining what to implement, while technical variability relates to product family engineering, defining how to implement it. A classification based on the dimensions of variability is proposed by Pohl *et al.* [112]: beyond **variability in time** (existence of different versions of an artifact that are valid at different times) and **variability in space** (existence of an artifact in different shapes at the same time) Pohl *et al.* claim that variability is important to different stakeholders and thus has different levels of visibility: **external variability** is visible to the customers while **internal variability**, that of domain artifacts, is hidden from them. Other classification proposals come from Meekel *et al.* [100] (feature, hardware platform, performances and attributes variability) or Bass *et al.* [68] who discuss about variability at the architectural level.

Central to the modeling of variability is the notion of *feature*, originally defined by Kang *et al.* as: *a prominent or distinctive user-visible aspect, quality or characteristic of a software system or systems* [96]. Based on this notion of *feature*, they proposed to use a *feature model* to model the variability in a SPL. A feature model consists of a *feature diagram* and other associated information: *constraints* and *dependency rules*. Feature diagrams provide a *graphical tree-like notation depicting the hierarchical organization of high level product functionalities* represented as features. The root of the tree refers to the complete system and is progressively decomposed into more refined features (tree nodes). Relations between nodes (features) are materialized by *decomposition edges* and *textual constraints*. Variability can be expressed in several ways. Presence or absence of a feature from a product is modeled using *mandatory* or *optional features*. Features are graphically represented as rectangles while some graphical elements (e.g., unfilled circle) are used to describe the variability (e.g., a feature may be optional).

Features can be organized into *feature groups*. Boolean operators *exclusive alternative (XOR)*, *inclusive alternative (OR)* or *inclusive (AND)* are used to select one, several or all the features from a feature group. Dependencies between features can be modeled using *textual constraints*: *requires* (presence of a feature requires the presence of another), *mutex* (presence of a feature automatically excludes another). Feature attributes can be also used for modeling quantitative (e.g., numerical) information. Constraints over attributes and features can be specified as well.

Modeling variability allows an organization to capture and select which version of which variant of any particular aspect is wanted in the system [77]. To implement it cheaply, quickly and safely, redoing by hand the tedious weaving of every aspect is not an option: some form of automation is needed to leverage the modeling of variability [72], [84]. Model Driven Engineering (MDE) makes it possible to automate this weaving process [95]. This requires that models are no longer informal, and that the weaving process is itself described as a program (which is as a matter of facts an executable meta-model [103]) manipulating these models to produce for instance a detailed design that can ultimately be transformed to code, or to test suites [111], or other software artifacts.

3.1.3. Component-based software development

Component-based software development [121] aims at providing reliable software architectures with a low cost of design. Components are now used routinely in many domains of software system designs: distributed systems, user interaction, product lines, embedded systems, etc. With respect to more traditional software artifacts (e.g., object oriented architectures), modern component models have the following distinctive features [83]: description of requirements on services required from the other components; indirect connections between components thanks to ports and connectors constructs [98]; hierarchical definition of components (assemblies of components can define new component types); connectors supporting various communication semantics [80]; quantitative properties on the services [75].

In recent years component-based architectures have evolved from static designs to dynamic, adaptive designs (e.g., SOFA [80], Palladio [73], Frascati [104]). Processes for building a system using a statically designed architecture are made of the following sequential lifecycle stages: requirements, modeling, implementation, packaging, deployment, system launch, system execution, system shutdown and system removal. If for any reason after design time architectural changes are needed after system launch (e.g., because requirements changed, or the implementation platform has evolved, etc) then the design process must be reexecuted from scratch (unless the changes are limited to parameter adjustment in the components deployed).

Dynamic designs allow for *on the fly* redesign of a component based system. A process for dynamic adaptation is able to reapply the design phases while the system is up and running, without stopping it (this is different from stop/redeploy/start). This kind of process supports *chosen adaptation*, when changes are planned and realized to maintain a good fit between the needs that the system must support and the way it supports them [97]. Dynamic component-based designs rely on a component meta-model that supports complex life cycles for components, connectors, service specification, etc. Advanced dynamic designs can also take platform changes into account at run-time, without human intervention, by adapting themselves [81], [124]. Platform changes and more generally environmental changes trigger *imposed adaptation*, when the system can no longer use its design to provide the services it must support. In order to support an eternal system [74], dynamic component based systems must separate architectural design and platform compatibility. This requires support for heterogeneity, since platform evolutions can be partial.

The Models@runtime paradigm denotes a model-driven approach aiming at taming the complexity of dynamic software systems. It basically pushes the idea of reflection one step further by considering the reflection layer as a real model "something simpler, safer or cheaper than reality to avoid the complexity, danger and irreversibility of reality [115]". In practice, component-based (and/or service-based) platforms offer reflection APIs that make it possible to introspect the system (which components and bindings are currently in place in the system) and dynamic adaptation (by applying CRUD operations on these components and bindings). While some of these platforms offer rollback mechanisms to recover after an erroneous adaptation, the idea of Models@runtime is to prevent the system from actually enacting an erroneous adaptation. In other words, the "model at run-time" is a reflection model that can be uncoupled (for reasoning, validation, simulation purposes) and automatically resynchronized.

Heterogeneity is a key challenge for modern component based system. Until recently, component based techniques were designed to address a specific domain, such as embedded software for command and control, or distributed Web based service oriented architectures. The emergence of the Internet of Things paradigm calls for a unified approach in component based design techniques. By implementing an efficient separation of concern between platform independent architecture management and platform dependent implementations, *Models@runtime* is now established as a key technique to support dynamic component based designs. It provides DIVERSE with an essential foundation to explore an adaptation envelop at run-time.

Search Based Software Engineering [92] has been applied to various software engineering problems in order to support software developers in their daily work. The goal is to automatically explore a set of alternatives and assess their relevance with respect to the considered problem. These techniques have been applied to craft software architecture exhibiting high quality of services properties [89]. Multi Objectives Search based techniques [86] deal with optimization problem containing several (possibly conflicting) dimensions to optimize. These techniques provide DIVERSE with the scientific foundations for reasoning and efficiently exploring an envelope of software configurations at run-time.

3.1.4. Validation and verification

Validation and verification (V&V) theories and techniques provide the means to assess the validity of a software system with respect to a specific correctness envelop. As such, they form an essential element of DIVERSE's scientific background. In particular, we focus on model-based V&V in order to leverage the different models that specify the envelop at different moments of the software development lifecycle.

Model-based testing consists in analyzing a formal model of a system (*e.g.*, activity diagrams, which capture high-level requirements about the system, statecharts, which capture the expected behavior of a software module, or a feature model, which describes all possible variants of the system) in order to generate test cases that will be executed against the system. Model-based testing [123] mainly relies on model analysis, constraint solving [85] and search-based reasoning [99]. DIVERSE leverages in particular the applications of model-based testing in the context of highly-configurable systems and [125] interactive systems [101] as well as recent advances based on diversity for test cases selection [93].

Nowadays, it is possible to simulate various kinds of models. Existing tools range from industrial tools such as Simulink, Rhapsody or Telelogic to academic approaches like Omega [109], or Xholon [0]. All these simulation environments operate on homogeneous environment models. However, to handle diversity in software systems, we also leverage recent advances in heterogeneous simulation. Ptolemy [79] proposes a common abstract syntax, which represents the description of the model structure. These elements can be decorated using different directors that reflect the application of a specific model of computation on the model element. Metropolis [69] provides modeling elements amenable to semantically equivalent mathematical models. Metropolis offers a precise semantics flexible enough to support different models of computation. ModHel'X [91] studies the composition of multi-paradigm models relying on different models of computation.

Model-based testing and simulation are complemented by runtime fault-tolerance through the automatic generation of software variants that can run in parallel, to tackle the open nature of software-intensive systems. The foundations in this case are the seminal work about N-version programming [67], recovery blocks [113] and code randomization [71], which demonstrated the central role of diversity in software to ensure runtime resilience of complex systems. Such techniques rely on truly diverse software solutions in order to provide systems with the ability to react to events, which could not be predicted at design time and checked through testing or simulation.

3.1.5. Empirical software engineering

The rigorous, scientific evaluation of DIVERSE's contributions is an essential aspect of our research methodology. In addition to theoretical validation through formal analysis or complexity estimation, we also aim at applying state-of-the-art methodologies and principles of empirical software engineering. This approach encompasses a set of techniques for the sound validation contributions in the field of software engineering, ranging from statistically sound comparisons of techniques and large-scale data analysis to interviews and systematic literature reviews [118], [116]. Such methods have been used for example to understand the impact of new software development paradigms [78]. Experimental design and statistical tests represent another major aspect of empirical software engineering. Addressing large-scale software engineering problems often requires the application of heuristics, and it is important to understand their effects through sound statistical analyses [66].

3.2. Research axis

Figure 1 illustrates the four dimensions of software diversity, which form the core research axis of DIVERSE: the **diversity of languages** used by the stakeholders involved in the construction of these systems; the **diversity of features** required by the different customers; the **diversity of runtime environments** in which software has to run and adapt; the **diversity of implementations** that are necessary for resilience through redundancy. These four axis share and leverage the scientific and technological results developed in the area of model-driven engineering in the last decade. This means that all our research activities are founded on sound abstractions to

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

reason about specific aspects of software systems, compose different perspectives and automatically generate parts of the system.

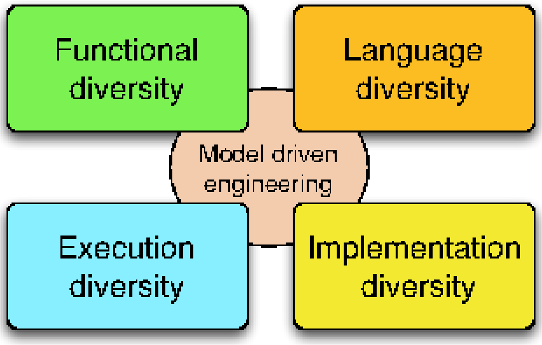


Figure 1. The four research axis of DIVERSE, which rely on a MDE scientific background

3.2.1. Software Language Engineering

The engineering of systems involves many different stakeholders, each with their own domain of expertise. Hence more and more organizations are adopting Domain Specific Modeling Languages (DSMLs) to allow domain experts to express solutions directly in terms of relevant domain concepts [117], [88]. This new trend raises new challenges about designing DSMLs, evolving a set of DSMLs and coordinating the use of multiple DSLs for both DSL designers and DSL users.

3.2.1.1. Challenges

Reusability of software artifacts is a central notion that has been thoroughly studied and used by both academics and industrials since the early days of software construction. Essentially, designing reusable artifacts allows the construction of large systems from smaller parts that have been separately developed and validated, thus reducing the development costs by capitalizing on previous engineering efforts. However, it is still hardly possible for language designers to design typical language artifacts (e.g. language constructs, grammars, editors or compilers) in a reusable way. The current state of the practice usually prevents the reusability of language artifacts from one language to another, consequently hindering the emergence of real engineering techniques around software languages. Conversely, concepts and mechanisms that enable artifacts reusability abound in the software engineering community.

Variability in modeling languages occur in the definition of the abstract and concrete syntax as well as in the specification of the language's semantics. The major challenges met when addressing the need for variability are: (i) set principles for modeling language units that support the modular specification of a modeling language; and (ii) design mechanisms to assemble these units in a complete language, according to the set of authorized variation points for the modeling language family.

A new generation of complex software-intensive systems (for example smart health support, smart grid, building energy management, and intelligent transportation systems) presents new opportunities for leveraging modeling languages. The development of these systems requires expertise in diverse domains. Consequently, different types of stakeholders (e.g., scientists, engineers and end-users) must work in a coordinated manner on various aspects of the system across multiple development phases. DSMLs can be used to support the work of domain experts who focus on a specific system aspect, but they can also provide the means for coordinating work across teams specializing in different aspects and across development phases. The support and integration of DSMLs leads to what we call **the globalization of modeling languages**, *i.e.* the use of multiple languages for the coordinated development of diverse aspects of a system. One can make an analogy with world globalization in which relationships are established between sovereign countries to regulate interactions (e.g., travel and commerce related interactions) while preserving each country's independent existence.

3.2.1.2. Scientific objectives

We address reuse and variability challenges through the investigation of the time-honored concepts of substitutability, inheritance and components, evaluate their relevance for language designers and provide tools and methods for their inclusion in software language engineering. We will develop novel techniques for the modular construction of language extensions with the support of model syntactical variability. From the semantics perspective, we investigate extension mechanisms for the specification of variability in operational semantics, focusing on static introduction and heterogeneous models of computation. The definition of variation points for the three aspects of the language definition provides the foundations for the novel concept Language Unit (LU) as well as suitable mechanisms to compose such units.

We explore the necessary breakthrough in software languages to support modeling and simulation of heterogeneous and open systems. This work relies on the specification of executable domain specific modeling languages (DSMLs) to formalize the various concerns of a software-intensive system, and of models of computation (MoCs) to explicitly model the concurrency, time and communication of such DSMLs. We develop a framework that integrates the necessary foundations and facilities for designing and implementing executable and concurrent domain-specific modeling languages. It also provides unique features to specify composition operators between (possibly heterogeneous) DSMLs. Such specifications are amenable to support the edition, execution, graphical animation and analysis of heterogeneous models. The objective is to provide both a significant improvement of MoCs and DSMLs design and implementation; and the simulation based validation and verification of complex systems.

We see an opportunity for the automatic diversification of programs' computation semantics, for example through the diversification of compilers or virtual machines. The main impact of this artificial diversity is to provide flexible computation and thus ease adaptation to different execution conditions. A combination of static and dynamic analysis could support the identification of what we call *plastic computation zones* in the code. We identify different categories of such zones: (i) areas in the code in which the order of computation can vary (e.g., the order in which a block of sequential statements is executed); (ii) areas that can be removed, keeping the essential functionality [119] (e.g., skip some loop iterations); (iii) areas that can replaced by alternative code (e.g., replace a try-catch by a return statement). Once we know which zones in the code can be randomized, it is necessary to modify the model of computation to leverage the computation plasticity. This consists in introducing variation points in the interpreter to reflect the diversity of models of computation. Then, the choice of a given variation is performed randomly at run-time.

3.2.2. Variability Modeling and Engineering

The systematic modeling of variability in software systems has emerged as an effective approach to document and reason about software evolutions and heterogeneity (*cf.* Section 3.1.2). Variability modeling characterizes an "envelope" of possible software variations. The industrial use of variability models and their relation to software artifact models require a complete engineering framework, including composition, decomposition, analysis, configuration and artifact derivation, refactoring, re-engineering, extraction, and testing. This framework can be used both to tame imposed diversity and to manage chosen diversity.

3.2.2.1. Challenges

A fundamental problem is that the **number of variants** can be exponential in the number of options (features). Already with 300 boolean configuration options, approximately 10^{90} configurations exist – more than estimated count of atoms in the universe. Domains like automotive or operating systems have to manage more than 10000 options (e.g., Linux). Practitioners face the challenge of developing billions of variants. It is easy to forget a necessary constraint, leading to the synthesis of unsafe variants, or to under-approximate the capabilities of the software platform. Scalable modelling techniques are therefore crucial to specify and reason about a very large set of variants.

Model-driven development supports two ways to deal with the increasing number of concerns in complex systems: (1) multi-view modeling, *i.e.* when modeling each concern separately, and variability modeling. However, there is little support to combine both approaches consistently. Techniques to integrate both approaches will enable the construction of a consistent set of views and variation points in each view.

The design, construction and maintenance of software families have a major impact on **software testing**. Among the existing challenges, we can cite: the selection of test cases for a specific variant; the evolution of test suites with integration of new variants; the combinatorial explosion of the number of software configurations to be tested. Novel model-based techniques for test generation and test management in a software product line context are needed to overcome state-of-the-art limits we already observed in some projects.

3.2.2.2. Scientific objectives

We aim at developing scalable techniques to automatically analyze variability models and their interactions with other views on the software intensive system (requirements, architecture, design). These techniques provide two major advancements in the state of the art: (1) an extension of the semantics of variability models in order to enable the definition of attributes (*e.g.*, cost, quality of service, effort) on features and to include these attributes in the reasoning; (2) an assessment of the consistent specification of variability models with respect to system views (since variability is orthogonal to system modeling, it is currently possible to specify the different models in ways that are semantically meaningless). The former aspect of analysis is tackled through constraint solving and finite-domain constraint programming, while the latter aspect is investigated through automatic search-based techniques (similar to genetic algorithms) for the exploration of the space of interaction between variability and view models.

We aim to develop procedures to reverse engineer dependencies and features' sets from existing software artefacts – be it source code, configuration files, spreadsheets (e.g., product comparison matrices) or requirements. We expect to scale up (e.g., for extracting a very large number of variation points) and guarantee some properties (e.g., soundness of configuration semantics, understandability of ontological semantics). For instance, when building complex software-intensive systems, textual requirements are captured in very large quantities of documents. In this context, adequate models to formalize the organization of requirements documents and automated techniques to support impact analysis (in case of changes in the requirements) have to be developed.

We aim at developing sound methods and tools to integrate variability management in model-based testing activities. In particular, we will leverage requirement models as an essential asset to establish formal relations between variation points and test models. These relations will form the basis for novel algorithms that drive the systematic selection of test configurations that satisfy well-defined test adequacy criteria as well as the generation of test cases for a specific product in the product line.

3.2.3. Heterogeneous and dynamic software architectures

Flexible yet dependable systems have to cope with heterogeneous hardware execution platforms ranging from smart sensors to huge computation infrastructures and data centers. Evolutions range from a mere change in the system configuration to a major architectural redesign, for instance to support addition of new features or a change in the platform architecture (new hardware is made available, a running system switches to low bandwidth wireless communication, a computation node battery is running low, etc). In this context, we need to devise formalisms to reason about the impact of an evolution and about the transition from one configuration to another. It must be noted that this axis focuses on the use of models to drive the evolution from design time to run-time. Models will be used to (i) systematically define predictable configurations and variation points through which the system will evolve; (ii) develop behaviors necessary to handle unpredicted evolutions.

3.2.3.1. Challenges

The main challenge is to provide new homogeneous architectural modelling languages and efficient techniques that enable continuous software reconfiguration to react to changes. This work handles the challenges of handling the diversity of runtime infrastructures and managing the cooperation between different stakeholders. More specifically, the research developed in this axis targets the following dimensions of software diversity.

Platform architectural heterogeneity induces a first dimension of imposed diversity (type diversity). Platform reconfigurations driven by changing resources define another dimension of diversity (deployment diversity). To deal with these imposed diversity problems, we will rely on model based runtime support for adaptation, in the spirit of the dynamic distributed component framework developed by the Triskell team. Since the runtime environment composed of distributed, resource constrained hardware nodes cannot afford the overhead of traditional runtime adaptation techniques, we investigate the design of novel solutions relying on models@runtime and on specialized tiny virtual machines to offer resource provisioning and dynamic reconfigurations. In the next two years this research will be supported by the InfraJVM project.

Diversity can also be an asset to optimize software architecture. Architecture models must integrate multiple concerns in order to properly manage the deployment of software components over a physical platform. However, these concerns can contradict each other (*e.g.*, accuracy and energy). In this context, we investigate automatic solutions to explore the set of possible architecture models and to establish valid trade-offs between all concerns in case of changes.

3.2.3.2. Scientific objectives

Automatic synthesis of optimal software architectures. Implementing a service over a distributed platform (*e.g.*, a pervasive system or a cloud platform) consists in deploying multiple software components over distributed computation nodes. We aim at designing search-based solutions to (i) assist the software architect in establishing a good initial architecture (that balances between different factors such as cost of the nodes, latency, fault tolerance) and to automatically update the architecture when the environment or the system itself change. The choice of search-based techniques is motivated by the very large number of possible software deployment architectures that can be investigated and that all provide different trade-offs between qualitative factors. Another essential aspect that is supported by multi-objective search is to explore different architectural solutions that are not necessarily comparable. This is important when the qualitative factors are orthogonal to each other, such as security and usability for example.

Flexible software architecture for testing and data management. As the number of platforms on which software runs increases and different software versions coexist, the demand for testing environments also increases. For example, to test a software patch or upgrade, the number of testing environments is the product of the number of running environments the software supports and the number of coexisting versions of the software. Based on our first experiment on the synthesis of cloud environment using architectural models, our objective is to define a set of domain specific languages to catch the requirement and to design cloud environments for testing and data management of future internet systems from data centers to things. These languages will be interpreted to support dynamic synthesis and reconfiguration of a testing environment.

Runtime support for heterogeneous environments. Execution environments must provide a way to account or reserve resources for applications. However, current execution environments such as the Java Virtual Machine do not clearly define a notion of application: each framework has its own definition. For example, in OSGi, an application is a component, in JEE, an application is most of the time associated to a class loader, in the Multi-Tasking Virtual machine, an application is a process. The challenge consists in defining an execution environment that provides direct control over resources (CPU, Memory, Network I/O) independently from the definition of an application. We propose to define abstract resource containers to account and reserve resources on a distributed network of heterogeneous devices.

3.2.4. Diverse implementations for resilience

Open software-intensive systems have to evolve over their lifetime in response to changes in their environment. Yet, most verification techniques assume a closed environment or the ability to predict all changes. Dynamic changes and evolutions thus represent a major challenge for these techniques that aim at assessing the correctness and robustness of the system. On the one hand, DIVERSE will adapt V&V techniques to handle diversity imposed by the requirements and the execution environment, on the other hand we leverage diversity to increase the robustness of software in face of unpredicted situations. More specifically, we address the following V&V challenges.

3.2.4.1. Challenges

One major challenge to build flexible and open yet dependable systems is that current software engineering techniques require architects to foresee all possible situations the system will have to face. However, openness and flexibility also mean unpredictability: unpredictable bugs, attacks, environmental evolutions, etc. Current fault-tolerance [113] and security [87] techniques provide software systems with the capacity of detecting accidental and deliberate faults. However, existing solutions assume that the set of bugs or vulnerabilities in a system does not evolve. This assumption does not hold for open systems, thus it is essential to revisit fault-tolerance and security solutions to account for diverse and unpredictable faults.

Diversity is known to be a major asset for the robustness of large, open, and complex systems (*e.g.*, economical or ecological systems). Following this observation, the software engineering literature provides a rich set of work that choose to implement diversity in software systems in order to improve robustness to attacks or to changes in quality of service. These works range from N-version programming to obfuscation of data structures or control flow, to randomization of instruction sets. An essential remaining challenge is to support the automatic synthesis and evolution of software diversity in open software-intensive systems. There is an opportunity to further enhance these techniques in order to cope with a wider diversity of faults, by multiplying the levels of diversity in the different software layers that are found in software-intensive systems (system, libraries, frameworks, application). This increased diversity must be based on artificial program transformations and code synthesis, which increase the chances of exploring novel solutions, better fitted at one point in time. The biological analogy also indicates that diversity should emerge as a side-effect of evolution, to prevent over-specialization towards one kind of diversity.

3.2.4.2. Scientific objectives

The main objective is to address one of the main limitations of N-version programming for fault-tolerant systems: the manual production and management of software diversity. Through automated injection of artificial diversity we aim at systematically increasing failure diversity and thus increasing the chances of early error detection at run-time. A fundamental assumption for this work is that software-intensive systems can be "good enough" [114], [126].

Proactive program diversification. We aim at establishing novel principles and techniques that favor the emergence of multiple forms of software diversity in software-intensive systems, in conjunction with the software adaptation mechanisms that leverage this diversity. The main expected outcome is a set of meta-design principles that maintain diversity in systems and the experimental demonstration of the effects of software diversity on the adaptive capacities of CASs. Higher levels of diversity in the system provide a pool of software solutions that can eventually be used to adapt to situations unforeseen at design time (bugs, crash, attacks, etc.). Principles of automated software diversification rely on the automated synthesis of variants in a software product line, as well as finer-grained program synthesis combining unsound transformations and genetic programming to explore the space of mutational robustness.

Multi-tier software diversification. We call multi-tier diversification the fact of diversifying several application software components simultaneously. The novelty of our proposal, with respect to the software diversity state of the art, is to diversify the application-level code (for example, diversify the business logics of the application), focusing on the technical layers found in web applications. The diversification of application software code is expected to provide a diversity of failures and vulnerabilities in web server deployment. Web server deployment usually adopts a form of the Reactor architecture pattern, for scalability purposes: multiple copies of the server software stack, called request handlers, are deployed behind a load balancer. This architecture is very favorable for diversification, since by using the multiplicity of request handlers running in a web server we can simultaneously deploy multiple combinations of diverse software components. Then, if one handler is hacked or crashes the others should still be able to process client requests.

4. Application Domains

4.1. From Embedded Systems to Service Oriented Architectures

From small embedded systems such as home automation products or automotive systems to medium sized systems such as medical equipment, office equipment, household appliances, smart phones; up to large Service Oriented Architectures (SOA), building a new application from scratch is no longer possible. Such applications reside in (group of) machines that are expected to run continuously for years without unrecoverable errors. Special care has then to be taken to design and validate embedded software, making the appropriate trade-off between various extra-functional properties such as reliability, timeliness, safety and security but also development and production cost, including resource usage of processor, memory, bandwidth, power, etc.

Leveraging ongoing advances in hardware, embedded software is playing an evermore crucial role in our society, bound to increase even more when embedded systems get interconnected to deliver ubiquitous SOA. For this reason, embedded software has been growing in size and complexity at an exponential rate for the past 20 years, pleading for a component based approach to embedded software development. There is a real need for flexible solutions allowing to deal at the same time with a wide range of needs (product lines modeling and methodologies for managing them), while preserving quality and reducing the time to market (such as derivation and validation tools).

We believe that building flexible, reliable and efficient embedded software will be achieved by reducing the gap between executable programs, their models, and the platform on which they execute, and by developing new composition mechanisms as well as transformation techniques with a sound formal basis for mapping between the different levels.

Reliability is an essential requirement in a context where a huge number of softwares (and sometimes several versions of the same program) may coexist in a large system. On one hand, software should be able to evolve very fast, as new features or services are frequently added to existing ones, but on the other hand, the occurrence of a fault in a system can be very costly, and time consuming. While we think that formal methods may help solving this kind of problems, we develop approaches where they are kept “behind the scene” in a global process taking into account constraints and objectives coming from user requirements.

Software testing is another aspect of reliable development. Testing activities mostly consist in trying to exhibit cases where a system implementation does not conform to its specifications. Whatever the efforts spent for development, this phase is of real importance to raise the confidence level in the fact that a system behaves properly in a complex environment. We also put a particular emphasis on on-line approaches, in which test and observation are dynamically computed during execution.

5. Highlights of the Year

5.1. Highlights of the Year

H2020 project accepted in the call ICT-10-2016 ’Software Technologies’, as coordinator.

The book “Engineering Modeling Languages” has been published by CRC Press. This book, co-authored by Benoit Combemale, Robert B. France, Jean-Marc Jézéquel, Bernhard Rumpe, Didier Vojtisek and Jim Steel, is the result of our respective expertise in model-driven engineering and software language engineering.

5.1.1. Awards

Silver Medal of the CNRS for Jean-Marc Jézéquel.

Second position for the ACM Student Research Competition: Thomas Degueule.

6. New Software and Platforms

6.1. FAMILIAR

KEYWORDS: Software line product - Configators - Customisation

SCIENTIFIC DESCRIPTION

FAMILIAR (for FeAture Model scrIpt Language for manIpulation and Automatic Reasoning) is a language for importing, exporting, composing, decomposing, editing, configuring, computing "diffs", refactoring, reverse engineering, testing, and reasoning about (multiple) feature models. All these operations can be combined to realize complex variability management tasks. A comprehensive environment is proposed as well as integration facilities with the Java ecosystem.

FUNCTIONAL DESCRIPTION

Familiar is an environment for large-scale product customisation. From a model of product features (options, parameters, etc.), Familiar can automatically generate several million variants. These variants can take many forms: software, a graphical interface, a video sequence or even a manufactured product (3D printing). Familiar is particularly well suited for developing web configurators (for ordering customised products online), for providing online comparison tools and also for engineering any family of embedded or software-based products.

- Participants: Mathieu Acher, Guillaume Becan, Olivier Barais
- Contact: Mathieu Acher
- URL: http://familiar-project.github.com

6.2. GEMOC Studio

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

SCIENTIFIC DESCRIPTION

The language workbench put together the following tools seamlessly integrated to the Eclipse Modeling Framework (EMF):

- Melange, a tool-supported meta-language to modularly define executable modeling languages with execution functions and data, and to extend (EMF-based) existing modeling languages.
- MoCCML, a tool-supported meta-language dedicated to the specification of a Model of Concurrency and Communication (MoCC) and its mapping to a specific abstract syntax and associated execution functions of a modeling language.
- GEL, a tool-supported meta-language dedicated to the specification of the protocol between the execution functions and the MoCC to support the feedback of the data as well as the callback of other expected execution functions.
- BCOoL, a tool-supported meta-language dedicated to the specification of language coordination patterns to automatically coordinates the execution of, possibly heterogeneous, models.
- Sirius Animator, an extension to the model editor designer Sirius to create graphical animators for executable modeling languages.

FUNCTIONAL DESCRIPTION

The GEMOC Studio is an eclipse package that contains components supporting the GEMOC methodology for building and composing executable Domain-Specific Modeling Languages (DSMLs). It includes the two workbenches: The GEMOC Language Workbench: intended to be used by language designers (aka domain experts), it allows to build and compose new executable DSMLs. The GEMOC Modeling Workbench: intended to be used by domain designersto create, execute and coordinate models conforming to executable DSMLs. The different concerns of a DSML, as defined with the tools of the language workbench, are automatically deployed into the modeling workbench. They parametrize a generic execution framework that provide various generic services such as graphical animation, debugging tools, trace and event managers, timeline, etc.

- Participants: Benoit Combemale, Dorian Leroy, Thomas Degueule, Erwan Bousse, Fabien Coulon and Didier Vojtisek
- Contact: Benoit Combemale
- URL: http://gemoc.org

6.3. Kevoree

Kevoree Core

KEYWORDS: Cloud - Deployment - Embedded - Domotique - Heterogeneity - Software Components - Architecture - Software component - Dynamic adaptation - M2M - Dynamic deployment

SCIENTIFIC DESCRIPTION

Kevoree is an open-source models@runtime platform (http://www.kevoree.org) to properly support the dynamic adaptation of distributed systems. Models@runtime basically pushes the idea of reflection [132] one step further by considering the reflection layer as a real model that can be uncoupled from the running architecture (e.g. for reasoning, validation, and simulation purposes) and later automatically resynchronized with its running instance.

Kevoree has been influenced by previous work that we carried out in the DiVA project [132] and the Entimid project [135] . With Kevoree we push our vision of models@runtime [131] farther. In particular, Kevoree provides a proper support for distributed models@runtime. To this aim we introduced the Node concept to model the infrastructure topology and the Group concept to model semantics of inter node communication during synchronization of the reflection model among nodes. Kevoree includes a Channel concept to allow for multiple communication semantics between remoteComponents deployed on heterogeneous nodes. All Kevoree concepts (Component, Channel, Node, Group) obey the object type design pattern to separate deployment artifacts from running artifacts. Kevoree supports multiple kinds of very different execution node technology (e.g. Java, Android, MiniCloud, FreeBSD, Arduino, ...).

Kevoree is distributed under the terms of the LGPL open source license.

Main competitors:

the Fractal/Frascati eco-system (http://frascati.ow2.org).

SpringSource Dynamic Module (http://spring.io/)

GCM-Proactive (http://proactive.inria.fr/)

OSGi (http://www.osgi.org)

Chef

Vagran (http://vagrantup.com/)

Main innovative features:

distributed models@runtime platform (with a distributed reflection model and an extensible models@runtime dissemination set of strategies).

Support for heterogeneous node type (from Cyber Physical System with few resources until cloud computing infrastructure).

Fully automated provisioning model to correctly deploy software modules and their dependencies.

Communication and concurrency access between software modules expressed at the model level (not in the module implementation).

Impact:

Several tutorials and courses have been performed this year at EJCP for French PhD student, at ECNU summer school for 82 chineese PhD students. See also the web page http://www.kevoree.org .

In 2015, we mainly created a new implementation in C# and we created an implementation for system containers for driving resources using Kevoree. We also use Kevoree in the context of Mohammed's PhD to create testing infrastructure on-demand.

FUNCTIONAL DESCRIPTION

Kevoree is an open-source models@runtime platform to properly support the dynamic adaptation of distributed systems. Models@runtime basically pushes the idea of reflection one step further by considering the reflection layer as a real model that can be uncoupled from the running architecture (e.g. for reasoning, validation, and simulation purposes) and later automatically resynchronized with its running instance.

- Participants: Jean Emile Dartois, Olivier Barais, Aymeric Hervieu, Johann Bourcier, Noel Plouzeau, Benoit Baudry, Maxime Tricoire, Jacky Bourgeois, Inti Gonzalez Herrera, Ivan Paez Anaya, Francisco Javier Acosta Padilla, Mohamed Boussaa and Manuel Leduc
- Partner: Université de Rennes 1
- Contact: Olivier Barais
- URL: http://kevoree.org/

6.4. Melange

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

SCIENTIFIC DESCRIPTION

Melange is a follow-up of the executable metamodeling language Kermeta, which provides a tool-supported dedicated meta-language to safely assemble language modules, customize them and produce new DSMLs. Melange provides specific constructs to assemble together various abstract syntax and operational semantics artifacts into a DSML. DSMLs can then be used as first class entities to be reused, extended, restricted or adapted into other DSMLs. Melange relies on a particular model-oriented type system that provides model polymorphism and language substitutability, i.e. the possibility to manipulate a model through different interfaces and to define generic transformations that can be invoked on models written using different DSLs. Newly produced DSMLs are correct by construction, ready for production (i.e., the result can be deployed and used as-is), and reusable in a new assembly.

Melange is tightly integrated with the Eclipse Modeling Framework ecosystem and relies on the meta-language Ecore for the definition of the abstract syntax of DSLs. Executable meta-modeling is supported by weaving operational semantics defined with Xtend. Designers can thus easily design an interpreter for their DSL in a non-intrusive way. Melange is bundled as a set of Eclipse plug-ins.

FUNCTIONAL DESCRIPTION

Melange is a language workbench which helps language engineers to mashup their various language concerns as language design choices, to manage their variability, and support their reuse. It provides a modular and reusable approach for customizing, assembling and integrating DSMLs specifications and implementations.

- Participants: Thomas Degueule, Benoit Combemale, Dorian Leroy, Erwan Bousse, Didier Vojtisek, Fabien Coulon, Jean-Marc Jezequel, Arnaud Blouin, Olivier Barais and David Mendez Acuna
- Contact: Benoit Combemale
- URL: http://melange-lang.org

6.5. Opencompare

KEYWORD: Software Product Line, Variability, MDE, Meta model, Configuration

FUNCTIONAL DESCRIPTION

Product comparison matrices (PCMs) are tabular data: supported and unsupported features are documented for both describing the product itself and for discriminating one product compared to another. PCMs abound – we are all using PCMs – and constitute a rich source of knowledge for easily comparing and choosing product. Yet the current practice is suboptimal both for humans and computers, mainly due to unclear semantics, heterogeneous forms of data, and lack of dedicated support.

OpenCompare.org is an ambitious project for the collaborative edition, the sharing, the standardisation, and the open exploitation of PCMs. The goal of OpenCompare.org is to provide an integrated set of tools (e.g., APIs, visualizations, configurators, editors) for democratizing their creation, import, maintenance, and exploitation.

- Participants: Mathieu Acher, Guillaume Becan and Sana Ben Nasr
- Contact: Mathieu Acher
- URL: http://opencompare.org

6.6. amiunique

KEYWORDS: Privacy - Browser fingerprinting

SCIENTIFIC DESCRIPTION

The amiunique web site has been deployed in the context of the DiverSE's research activities on browser fingerprinting and how software diversity can be leveraged in order to mitigate the impact of fingerprinting on the privacy of users. The construction of a dataset of genuine fingerprints is essential to understand in details how browser fingerprints can serve as unique identifiers and hence what should be modified in order to mitigate its impact privacy. This dataset also supports the large-scale investigation of the impact of web technology advances on fingerprinting. For example, we can analyze in details the impact of the HTML5 canvas element or the behavior of fingerprinting on mobile devices.

The whole source code of amiunique is open source and is distributed under the terms of the MIT license.

Similar sites: Panopticlick https://panopticlick.eff.org/ BrowserSpy http://browserspy.dk/http://noc.to/ Main innovative features: canvas fingerprinting WebGL fingerprinting advanced JS features (platform, DNT, etc.)

Impact: The website has been showcased in several professional forums in 2014 and 2015 (Open World Forum 2014, FOSSA'14, FIC'15, ICT'15) and it has been visited by more than 100000 unique visitors in one year.

FUNCTIONAL DESCRIPTION

This web site aims at informing visitors about browser fingerprinting and possible tools to mitigate its effect, as well as at collecting data about the fingerprints that can be found on the web. It collects browser fingerprints with the explicit agreement of the users (they have to click on a button on the home page). Fingerprints are composed of 17 attributes, which include regular HTTP headers as well as the most recent state of the art techniques (canvas fingerprinting, WebGL information).

- Participants: Benoit Baudry and Pierre Laperdrix
- Partner: INSA Rennes
- Contact: Benoit Baudry
- URL: https://amiunique.org/

7. New Results

7.1. Results on Variability modeling and management

7.1.1. Feature Model Synthesis: Algorithms and Empirical Studies

We attack the problem of synthesising feature models by considering both configuration semantics and ontological semantics of a feature model. We define a generic synthesis procedure that computes the likely siblings or parent candidates for a given feature. We develop six heuristics for clustering and weighting the logical, syntactical and semantical relationships between feature names. We then perform an empirical evaluation on hundreds of feature models, coming from the SPLOT repository and Wikipedia. We provide evidence that a fully automated synthesis (i.e., without any user intervention) is likely to produce models far from the ground truths. As the role of the user is crucial, we empirically analyze the strengths and weaknesses of heuristics for computing ranking lists and different kinds of clusters. We show that a hybrid approach mixing logical and ontological techniques outperforms state-of-the-art solutions.

Numerous synthesis techniques and tools have been proposed, but only a few consider both configuration and ontological semantics of a feature model. We also boil down several feature model management operations to a synthesis problem. Our approach, the FAMILIAR environment, and empirical results support researchers and practitioners working on feature models. The synthesis problem is a core issue when reverse engineering, merging, slicing, or refactoring feature models. An article has been published in 2016 at Empirical Software Engineering journal, a major avenue for software engineering research [19].

7.1.2. Product Comparison Matrix

Product Comparison Matrices (PCMs) are widely used for documenting or comparing a set of products. PCMs are simple tabular data in which products are usually organized as rows, features as columns, while each cell define how a product implements the corresponding feature. We develop metamodeling and feature modeling techniques for formalizing PCMs. We perform numerous empirical experiments with users, tools, and data for validating our proposal. We also develop automated techniques to extract PCMs out of informal product descriptions, written in natural language. We establish a connection between PCMs and variability modeling formalism, which is of interest for the product line community. OpenCompare is a direct output of this research and is an important step towards the creation of a community around PCMs. We mined millions of Wikipedia tabular data together with end-users and developers to cross-validate our model-based approach [19]. We also mined data from BestBuy [17].

7.1.3. Machine Learning and Variability Testing

We propose the use of a machine learning approach to infer variability constraints from an oracle that is able to assess whether a given configuration is correct. We propose an automated procedure to randomly generate configurations, classify them according to the oracle, and synthesize cross-tree constraints. We validate our approach on a product-line video generator, using a simple computer vision algorithm as an oracle. We show that an interesting set of cross-tree constraint can be generated, with reasonable precision and recall. Our learning-based testing technique complements our initial effort in engineering an industrial video generator. The use of learning allows to significantly narrow the configuration space and discover complex constraints, hard to discover even for experts. We conduct a series of work in the computer vision domain to generate variants of videos, investigating the usefulness and effectiveness of variability techniques in novel areas. Our approach is novel and general: the same principles can be applied to other configurable systems [55].

7.1.4. Enumeration of All Feature Model Configurations

Feature models are widely used to encode the configurations of a software product line in terms of mandatory, optional and exclusive features as well as propositional constraints over the features. Numerous computationally expensive procedures have been developed to model check, test, configure, debug, or compute relevant information of feature models. We explore the possible improvement of relying on the enumeration of all configurations when performing automated analysis operations. We tackle the challenge of how to scale the existing enumeration techniques by relying on distributed computing. We show that the use of distributed computing techniques might offer practical solutions to previously unsolvable problems and opens new perspectives for the automated analysis of software product lines [40].

7.1.5. Software Unbundling

Unbundling is a phenomenon that consists of dividing an existing software artifact into smaller ones. It can happen for different reasons, one of them is the fact that applications tend to grow in functionalities and sometimes this can negatively influence the user experience. It can be seen as a way to produce different variants of an application. For example, mobile applications from well-known companies are being divided into simpler and more focused new ones. Despite its current importance, little is known or studied about unbundling or about how it relates to existing software engineering approaches, such as modularization. Consequently, recent cases point out that it has been performed unsystematically and arbitrarily. Our main goal is to present this novel and relevant concept and its underlying challenges in the light of software engineering, also exemplifying it with recent cases. We relate unbundling to standard software modularization, presenting the new motivations behind it, the resulting problems, and drawing perspectives for future support in the area [23].

7.1.6. Featured Model Types

By analogy with software product reuse, the ability to reuse (meta)models and model transformations is key to achieve better quality and productivity. To this end, various opportunistic reuse techniques have been developed, such as higher-order transformations, metamodel adaptation, and model types. However, in contrast to software product development that has moved to systematic reuse by adopting (model-driven) software product lines, we are not quite there yet for modelling languages, missing economies of scope and automation opportunities. Our vision is to transpose the product line paradigm at the metamodel level, where reusable assets are formed by metamodel and transformation fragments and "products" are reusable language building blocks (model types). We introduce featured model types to concisely model variability amongst metamodelling elements, enabling configuration, automated analysis, and derivation of tailored model types [53].

7.1.7. A Formal Modeling and Analysis Framework for SPL of Pre-emptive Real-time Systems

We present a formal analysis framework to analyze a family of platform products w.r.t. real-time properties. First, we propose an extension of the widely-used feature model, called Property Feature Model (PFM), that distinguishes features and properties explicitly Second, we present formal behavioral models of components of a real-time scheduling unit such that all real-time scheduling units implied by a PFM are automatically composed to be analyzed against the properties given by the PFM. We apply our approach to the verification of the schedulability of a family of scheduling units using the symbolic and statistical model checkers of Uppaal [44].

7.1.8. Exploration of Architectural Variants

In systems engineering, practitioners shall explore numerous architectural alternatives until choosing the most adequate variant. The decision-making process is most of the time a manual, time-consuming, and error-prone activity. The exploration and justification of architectural solutions is ad-hoc and mainly consists in a series of tries and errors on the modeling assets.

We report on an industrial case study in which we apply variability modeling techniques to automate the assessment and comparison of several candidate architectures (variants). We first describe how we can use a model-based approach such as the Common Variability Language (CVL) to specify the architectural variability. We show that the selection of an architectural variant is a multi-criteria decision problem in which there are numerous interactions (veto, favor, complementary) between criteria. We present a tooled process for exploring architectural variants integrating both CVL and the MYRIAD method for assessing and comparing variants based on an explicit preference model coming from the elicitation of stakeholders' concerns. This solution allows understanding differences among variants and their satisfactions with respect to criteria. Beyond variant selection automation improvement, this experiment results highlight that the approach improves rationality in the assessment and provides decision arguments when selecting the preferred variants. It is a joint work and collaboration with Thales [47].

7.1.9. A Complexity Tale: Web Configurators

Online configurators are basically everywhere. From physical goods (cars, clothes) to services (cloud solutions, insurances, etc.) such configurators have pervaded many areas of everyday life, in order to provide the customers products tailored to their needs. Being sometimes the only interfaces between product suppliers and consumers, much care has been devoted to the HCI aspects of configurators, aiming at offering an enjoyable buying experience. However, at the backend, the management of numerous and complex configuration options results from ad-hoc process rather than a systematic variability-aware engineering approach. We present our experience in analysing web configurators and formalising configuration options in terms of feature models or product configuration matrices. We also consider behavioural issues and perspectives on their architectural design [32].

7.2. Results on Software Language Engineering

7.2.1. Safe Model Polymorphism for Flexible Modeling

Domain-Specific Languages (DSLs) are increasingly used by domain experts to handle various concerns in systems and software development. To support this trend, the Model-Driven Engineering (MDE) community has developed advanced techniques for designing new DSLs. However, the widespread use of independently developed, and constantly evolving DSLs is hampered by the rigidity imposed to the language users by the DSLs and their tooling, e.g., for manipulating a model through various similar DSLs or successive versions of a given DSL. In [24] we propose a disciplined approach that leverages type groups' polymorphism to provide an advanced type system for manipulating models, in a polymorphic way, through different DSL interfaces. A DSL interface, aka. model type, specifies a set of features, or services, available on the model it types, and subtyping relations among these model types define the safe substitutions. This type system complements the Melange language workbench and is seamlessly integrated into the Eclipse Modeling Framework (EMF), hence providing structural interoperability and compatibility of models between EMF-based tools. We illustrate the validity and practicability of our approach by bridging safe interoperability between different semantic and syntactic variation points of a finite-state machine (FSM) language, as well as between successive versions of the Unified Modeling Language (UML).

7.2.2. Execution Framework for Model Debugging

The development and evolution of an advanced modeling environment for a Domain-Specific Modeling Language (DSML) is a tedious task, which becomes recurrent with the increasing number of DSMLs involved in the development and management of complex software-intensive systems. Recent efforts in language workbenches result in advanced frameworks that automatically provide syntactic tooling such as advanced editors. However, defining the execution semantics of languages and their tooling remains mostly hand crafted. Similarly to editors that share code completion or syntax highlighting, the development of advanced debuggers, animators, and others execution analysis tools shares common facilities, which should be reused among various DSMLs. In [37] we present the execution framework offered by the GEMOC studio, an Eclipse-based language and modeling workbench. The framework provides a generic interface to plug in different execution engines associated to their specific metalanguages used to define the discrete-event operational semantics of DSMLs. It also integrates generic runtime services that are shared among the approaches used to implement the execution semantics, such as graphical animation or omniscient debugging.

7.2.3. Variability Management in Language Families

The use of domain-specific languages (DSLs) has become a successful technique in the development of complex systems. Nevertheless, the construction of this type of languages is time-consuming and requires highly-specialized knowledge and skills. An emerging practice to facilitate this task is to enable reuse through the definition of language modules which can be later put together to build up new DSLs. In [29], we report on an effort for organizing the literature on language product line engineering. More precisely, we propose a definition for the life-cycle of language product lines, and we use it to analyze the capabilities of current approaches. In addition, we provide a mapping between each approach and the technological space it supports.

Still, the identification and definition of language modules are complex and error-prone activities, thus hindering the reuse exploitation when developing DSLs. In [50], [51], we propose a computer-aided approach to i) identify potential reuse in a set of legacy DSLs; and ii) capitalize such potential reuse by extracting a set of reusable language modules with well defined interfaces that facilitate their assembly. We validate our approach by using realistic DSLs coming out from industrial case studies and obtained from public GitHub repositories. We also developed a publicly available tool, namely Puzzle, that uses static analysis to facilitate the detection of specification clones in DSLs implemented under the executable metamodeling paradigm. Puzzle also enables the extraction specification clones as reusable language modules that can be later used to build up new DSLs.

7.2.4. A Tool-Supported Approach for Concurrent Execution of Heterogeneous Models

In the software and systems modeling community, research on domain-specific modeling languages (DSMLs) is focused on providing technologies for developing languages and tools that allow domain experts to develop system solutions efficiently. Unfortunately, the current lack of support for explicitly relating concepts expressed in different DSMLs makes it very difficult for software and system engineers to reason about information spread across models describing different system aspects. As a particular challenge, we investigate in [38] relationships between, possibly heterogeneous, behavioral models to support their concurrent execution. This is achieved by following a modular executable metamodeling approach for behavioral semantics understanding, reuse, variability and composability. This approach supports an explicit model of concurrency (MoCC) and domain-specific actions (DSA) with a well-defined protocol between them (incl., mapping, feedback and callback) reified through explicit domain-specific events (DSE). The protocol is then used to infer a relevant behavioral language interface for specifying coordination patterns to be applied on conforming executable models. All the tooling of the approach is gathered in the GEMOC studio, and outlined in the next section. Currently, the approach is experienced on a systems engineering language provided by Thales, named Capella.

7.2.5. Various Dimensions of Reuse

Reuse, enabled by modularity and interfaces, is one of the most important concepts in software engineering. This is evidenced by an increasingly large number of reusable artifacts, ranging from small units such as classes to larger, more sophisticated units such as components, services, frameworks, software product lines, and concerns. We give evidence in [43] that a canonical set of reuse interfaces has emerged over time: the variation, customization, and usage interfaces (VCU). A reusable artifact that provides all three interfaces reaches the highest potential of reuse, as it explicitly exposes how the artifact can be manipulated during the reuse process along these three dimensions. We demonstrate the wide applicability of the VCU interfaces along two axes: across abstraction layers of a system specification and across existing reuse techniques. The former is shown with the help of a comprehensive case study including reusable requirements, software, and hardware models for the authorization domain. The latter is shown with a discussion on how the VCU interfaces relate to existing reuse techniques.

7.2.6. Modeling for Sustainability

The complex problems that computational science addresses are more and more benefiting from the progress of computing facilities (e.g., simulators, librairies, accessible languages). Nevertheless, the actual solutions call for several improvements. Among those, we address the needs for leveraging on knowledge and expertise by focusing on Domain-Specific Modeling Languages application. In this work we explored, through concrete experiments, how the last DSML research help getting closer the problem and implementation spaces.

Various disciplines use models for different purposes. While engineering models, including software engineering models, are often developed to guide the construction of a non- existent system, scientific models, in contrast, are created to better understand a natural phenomenon (i.e., an already existing system). An engineering model may incorporate scientific models to build a system. Both engineering and scientific models have been used to support sustainability, but largely in a loosely-coupled fashion, independently developed and maintained from each other. Due to the inherent complex nature of sustainability that must balance trade-offs between social, environmental, and economic concerns, modeling challenges abound for both the scientific and engineering disciplines. In [39] we propose a vision that synergistically combines engineering and scientific models to enable broader engagement of society for addressing sustainability concerns, informed decision-making based on more accessible scientific models and data, and automated feed-back to the engineering models to support dynamic adaptation of sustainability systems. To support this vision, we identify a number of challenges to be addressed with particular emphasis on the socio-technical benefits of modeling.

As first experiments, we presented at the EclipseCon France, Europe and North America 2016, an approach to develop smart cyber physical systems in charge of managing the production, distribution and consumption of energies (e.g., water, electricity). The main objective is to enable a broader engagement of society, while supporting a more informed decision-making, possibly automatically, on the development and run-time adaptation of sustainability systems (e.g., smart grid, home automation, smart cities). We illustrate this approach through a system that allows farmers to simulate and optimize their water consumption by combining the model of a farming system together with agronomical models (e.g., vegetable and animal lifecycle) and open data (e.g., climate series). To do so, we use Model Driven Engineering (MDE) and Domain Specific Languages (DSL) to develop such systems driven by scientific models that define the context (e.g., environment, social and economy), and model experiencing environments to engage general public and policy makers.

7.2.7. Formal Specification of a Packet Filtering Language Using the K Framework

Many project-specific languages, including in particular filtering languages, are defined using non-formal specifications written in natural languages. This leads to ambiguities and errors in the specification of those languages. In [46] we report on an industrial experiment on using a tool-supported language specification framework (K) for the formal specification of the syntax and semantics of a filtering language having a complexity similar to those of real-life projects. This experimentation aims at estimating, in a specific industrial setting, the difficulty and benefits of formally specifying a packet filtering language using a tool-supported formal approach.

7.2.8. Correct-by-construction model driven engineering composition operators

Model composition is a crucial activity in Model Driven Engineering both to reuse validated and verified model elements and to handle separately the various aspects in a complex system and then weave them while preserving their properties. Many research activities target this compositional validation and verification (V & V) strategy: allow the independent assessment of components and minimize the residual V & V activities at assembly time. However, there is a continuous and increasing need for the definition of new composition operators that allow the reconciliation of existing models to build new systems according to various requirements. These ones are usually built from scratch and must be systematically verified to assess that they preserve the properties of the assembled elements. This verification is usually tedious but is mandatory to avoid verifying the composite system for each use of the operators. Our work addresses these issues, we first target the use of proof assistants for specifying and verifying compositional verification frameworks relying on formal verification techniques instead of testing and proofreading. Then, using a divide and conquer approach, we focus on the development of elementary composition operators that are easy to verify and can be used to further define complex composition operators. In our approach [27], proofs for the complex operators are then obtained by assembling the proofs of the basic operators. To illustrate our proposal, we use the Coq proof assistant to formalize the language-independent elementary composition operators Union and Substitution and the proof that the conformance of models with respect to metamodels is preserved during composition. We show that more sophisticated composition operators that share parts of the implementation and have several properties in common (especially: aspect oriented modeling composition approach, invasive software composition, and package merge) can then be built from the basic ones, and that the proof of conformance preservation can also be built from the proofs of basic operators.

7.2.9. Engineering Modeling Languages

The DiverSE project-team is deeply involved in transferring research knowledge into education. In particular, one book in English have been published in 2016 as a textbook [59]. The book cover the broad scope of MDE, and are based on the experience of the project-team members.

7.3. Results on Heterogeneous and dynamic software architectures

We have selected three main contributions : two are in the field of runtime management, while the third one is in the field of non-functionnal software testing.

7.3.1. Precise and efficient resource management using models@runtime

Contribution. We have developed an efficient monitoring framework to quickly spot an abnormal resource consumption within a complex application. In these papers [25], we have proposed an optimistic adaptive monitoring system to determine the faulty components of an application. Suspected components are finely analyzed by the monitoring system, but only when required. Unsuspected components are left untouched and execute normally.

Originality. Current solutions that perform permanent and extensive monitoring to detect anomalies induce high overhead on the system, and can, by themselves, make the system unstable. Our system performs localized just-in-time monitoring that decreases the accumulated overhead of the monitoring system. Through our evaluation, we show that our technique correctly detects faulty components, while reducing overhead by 92.98 on average%.

Impact. Beyond the scientific originality of this work, the main impacts of this novel approach approach to monitor software component performance has been to (i) reinforce DIVERSE's visibility in the academic and industrial communities on software components and (ii) to create several research tracks that are currently explored in different projects of the team (HEADS and B-com PhD thesis). This work has been integrated within the Kevoree platform.

7.3.2. Dynamic web application using models@runtime

Contribution. We have developed a component-based platform supporting the development of dynamically adaptable single Web page applications. An important part of this contribution lies in the possibility to dynamically move code from the server to the client side allowing a great flexibility in the performance management. This contribution [56] is based on a models@runtime approach and has been implemented in our open source KevoreeJS platform.

Originality. Current solutions to create single Web page application are limited to a static code repartition between clients and server, thus limiting the flexibility at runtime.

Impact. Beyond the scientific originality of this work, the main impacts of this novel approach to monitor software component performance has been to (i) reinforce DIVERSE's visibility in the open-source community, (ii) to start several research tracks that are currently explored in different projects of the team (HEADS, STAMP, GRevis). This platforms is modular, one of the component has a monthly download count greater than 100k [0]).

7.3.3. Testing non-functional behavior of compiler and code generator

Contribution. We have developed NOTICE [36], [35], a component-based framework for non-functional testing of compilers through the monitoring of generated code in a controlled sand-boxing environment. In this work, we have proposed an automatic way of testing non-functional properties of compilers, while optimizing the generated application with respect to a set of specific non-functional properties (CPU, memory usage, energy consumption, *etc.*).

Originality. Compiler users generally apply different optimizations to generate efficient code with respect to specific non-functional properties such as energy consumption, execution time, etc. However, due to the huge number of optimizations provided by modern compilers, finding the best optimization sequence for a specific objective and a given program is more and more challenging.

Impact. Beyond the scientific originality of this work, the main impact of this novel approach is to enable the auto-tuning of compilers according to user requirements and to construct optimizations that yield to performance results that are better than standard optimization levels.

7.3.4. Automatic Microbenchmark Generation to Prevent Dead Code Elimination and Constant Folding

[0] https://www.npmjs.com/package/npmi

Contribution. Microbenchmarking consists of evaluating, in isolation, the performance of small code segments that play a critical role in large applications. The accuracy of a microbenchmark depends on two critical tasks: wrap the code segment into a payload that faithfully recreates the execution conditions that occur in the large application; build a scaffold that runs the payload a large number of times to get a statistical estimate of the execution time. While recent frameworks such as the Java Microbenchmark Harness (JMH) take care of the scaffold challenge, developers have very limited support to build a correct payload. This year, we focus on the automatic generation of pay-loads, starting from a code segment selected in a large application [54]. In particular, we aim at preventing two of the most common mistakes made in microbenchmarks: dead code elimination and constant folding. Since a microbench-mark is such a small program, if not designed carefully, it will be *over-optimized* by the JIT and result in distorted time measures. Our technique hence automatically extracts the segment into a compilable payload and generates additional code to prevent the risks of *over-optimization*. The whole approach is embedded in a tool called AutoJMH, which generates payloads for JMH scaffolds. We validate the capabilities AutoJMH, showing that the tool is able to process a large percentage of segments in real programs. We also show that AutoJMH can match the quality of payloads handwritten by performance experts and outperform those written by professional Java developers without experience un microbenchmarking.

7.3.5. Collaborations

This year, we had a close and fruitful collaboration with the industrial partners that are involved in the HEADS and Occiware projects, in particular an active interaction with the Tellu company in Norway in the Heads context [49]. Tellu relies on Kevoree and KevoreeJS to build their health management systems. They will be also a active member the new Stamp project led by DIVERSE. We can cite also an active collaboration with Orange Labs through Kevin Corre's joint PhD thesis. Another joint industrial (CIFRE) PhD started in September 2016, and we are also partner in a new starting FUI project. Finally, DIVERSE collaborates with the B-COM IRT (https://b-com.com/en), as one permanent member has a researcher position of one day per week at B-COM and a new joint PhD started in September [52].

At the academic level we collaborate actively with the Spiral team at Inria Lille (several joint projects), the Tacoma team (with two co-advised PhD students), the Myriad team (1 co-advised PhD student) and we have started two collaborations with the ASAP team.

7.4. Results on Diverse Implementations for Resilience

Diversity is acknowledged as a crucial element for resilience, sustainability and increased wealth in many domains such as sociology, economy and ecology. Yet, despite the large body of theoretical and experimental science that emphasizes the need to conserve high levels of diversity in complex systems, the limited amount of diversity in software-intensive systems is a major issue. This is particularly critical as these systems integrate multiple concerns, are connected to the physical world through multiple sensors, run eternally and are open to other services and to users. Here we present our latest observational and technical results about (i) new approaches to increase diversity in software systems, and (ii) software testing to assess the validity of software.

7.4.1. Software diversification

A main achievement in our investigations of software diversity, is a large scale analysis of browser fingerprints [45]. Browser fingerprinting consists in collecting information about a user's browser and its execution environment. A distinctive feature of these fingerprints is that they are unique and can be used to track users. We show that innovations in HTML5 provide access to highly discriminating attributes, notably with the use of the Canvas API which relies on multiple layers of the user's system. In addition, we show that browser fingerprinting is as effective on mobile devices as it is on desktops and laptops, albeit for radically different reasons due to their more constrained hardware and software environments. We also evaluate how browser fingerprinting could stop being a threat to user privacy if some technological evolutions continue (e.g., disappearance of plugins) or are embraced by browser vendors (e.g., standard HTTP headers).

As for automatic diversification of programs, we have had a strong focus on runtime transformations. Online Genetic Improvement embeds the ability to evolve and adapt inside a target software system enabling it to improve at runtime without any external dependencies or human intervention. We recently developed a general purpose tool enabling Online Genetic Improvement in software systems running on the java virtual machine. This tool, dubbed ECSELR, is embedded inside extant software systems at runtime, enabling such systems to autonomously generate diverse variants [31]. We have also worked on diversification against just-in-Time (JIT) Spraying: a technique that embeds return-oriented programming (ROP) gadgets in arithmetic or logical instructions as immediate offsets. We introduce libmask, a JIT compiler extension that transforms constants into global variables and marks the memory area for these global variables as read only. Hence, any constant is referred to by a memory address making exploitation of arithmetic and logical instructions more difficult. Then, these memory addresses are randomized to further harden the security [42].

7.4.2. Software testing

Our work in the area of software testing focuses on tailoring the testing tools (analysis, generation, oracle, etc.) to specific domains and purposes. This allows us to consider domain specific knowledge (e.g., architectural patterns for GUI implementation) in order to increase the relevance and the efficiency of testing. The main results of this year are about test case refactoring and testing code generators.

Software developers design test suites to verify that software meets its expected behaviors. Yet, many dynamic analysis techniques are performed on the exploitation of execution traces from test cases. In practice, one test case may imply various behaviors. However, the execution of a test case only yields one trace, which can hide the others. We have developed a new technique of test code refactoring, which splits a test case into small test fragments that cover a simpler part of the control flow to provide better support for dynamic analysis. This technique can effectively improve the execution traces of the test suite: exception contracts are better verified via applying this refactoring to original test suites [30].

Finding the smallest set of valid test configurations that ensure sufficient coverage of the system's feature interactions is essential, especially when the execution of test configurations is costly or time-consuming. However, this problem is NP-hard in general and approximation algorithms have often been used to address it in practice. We explore an approach based on constraint programming to increase the effectiveness of configuration testing while keeping the number of configurations as low as possible. For 79% of 224 feature models, our technique generated up to 60% fewer test configurations than the competitor tools [26].

The intensive use of generative programming techniques provides an elegant engineering solution to deal with the heterogeneity of platforms and technological stacks. Yet, producing correct and efficient code generator is complex and error-prone. We describe a practical approach based on a runtime monitoring infrastructure to automatically check the potential inefficient code generators. We evaluate our approach by analyzing the performance of Haxe, a popular high-level programming language that involves a set of cross-platform code generators. The results show that our approach is able to detect some performance inconsistencies that reveal real issues in Haxe code generators [36], [35]

Graphical User Interfaces (GUIs) intensively rely on event-driven programming: widgets send GUI events, which capture users' interactions, to dedicated objects called *controllers*. Controllers implement several *GUI listeners* that handle these events to produce GUI commands. We study to what extent the number of GUI commands that a GUI listener can produce has an impact on the code quality. We then identify a new type of design smell, called *Blob listener* that characterizes GUI listeners that can produce more than two GUI commands. We propose a systematic static code analysis procedure that searches for *Blob Listener* instances that we implement in *InspectorGuidget* [48].

8. Partnerships and Cooperations

8.1. National Initiatives

8.1.1. ANR

8.1.1.1. GEMOC

- Coordinator: Inria (DIVERSE)
- Other partners: ENSTA Bretagne, Inria, IRIT, I3S, Obeo, Thales
- Dates: 2012-2016
- Abstract: GEMOC focuses on a generic framework for heterogeneous software model execution and dynamic analysis. This work has the ambition to propose an innovative environment for the design of complex software-intensive systems by providing: a formal framework that integrates state-of-the-art in model-driven engineering (MDE) to build domain-specific modeling languages (DSMLs), and models of computation (MoC) to reason over the composition of heterogeneous concerns; an open-source design and modeling environment associated to a well-defined method for the definition of DSMLs, MoCs and rigorous composition of all concerns for execution and analysis purposes.

  This requires addressing two major scientific issues: the design and verification of a formal framework to combine several different DSMLs relying on distinct MoCs; the design and validation of a methodology for DSMLs and MoC development. GEMOC aims at participating in the development of next generation MDE environments through a rigorous, tool-supported process for the definition of executable DSMLs and the simulation of heterogeneous models.

8.1.1.2. SOPRANO

- Coordinator: CEA
- CEA, University of Paris-Sud, Inria Rennes, OcamlPro, Adacore
- Dates: 2014-2017
- Abstract: Today most major verification approaches rely on automatic external solvers. However these solvers do not fill the current and future needs for verification: lack of satisfying model generation, lack of reasoning on difficult theories (e.g. floating-point arithmetic), lack of extensibility for specific or new needs. The SOPRANO project aims at solving these problems and prepare the next generation of verification-oriented solvers by gathering experts from academia and industry. We will design a new framework for the cooperation of solvers, focused on model generation and borrowing principles from SMT (current standard) and CP (well-known in optimisation). These ideas will be implemented in an open-source platform, with regular evaluations from the industrial partners.

8.1.1.3. Gdiv MRSE

- Coordinator: B. Baudry
- Inria Rennes
- Dates: 2014-2016
- Abstract: The objective of GDiv is to setup a strong network of European partners around the core team composed of Inria and SINTEF. This network will gather another academic partner and between 3 and 5 industry partners in the areas of software development and deployment. The project proposal setup by the GDiv network will address the risks of large scale software reuse through integrated, multi-level software diversification techniques.

8.1.2. BGLE / LEOC

8.1.2.1. CONNEXION

- Coordinator: EDF
- Other partners: Atos WorldGrid, Rolls-Royce Civil Nuclear, Corys TESS, Esterel Technologies, All4Tec, Predict, CEA, Inria, CNRS / CRAN, ENS Cachan, LIG, Telecom ParisTech
- Dates: 2012-2016
- Abstract: The cluster CONNEXION (*digital command CONntrol for Nuclear EXport and renovatION*) aims to propose and validate an innovative architecture platforms suitable control systems for nuclear power plants in France and abroad. In this project the Triskell team investigates methods and tools to (i) automatically analyze and compare regulatory requirements evolutions and geographical differences; (ii) automatically generate test cases for critical interactive systems.

8.1.2.2. CLARITY

- Coordinator: Obéo
- Other partners: AIRBUS, Airbus Defence and Space, All4tec, ALTRAN Technologies, AREVA, Artal, C.E.S.A.M.E.S., Eclipse Foundation Europe, Inria Sophia Antipolis Méditerranée, PRFC, Scilab Enterprises, Thales Global Services, Thales Alenia Space, Thales Research & Technology, Thales Systèmes Aéroportés, Université de Rennes 1.
- Dates: 2014-2017
- Abstract: The CLARITY project aims to establish an international dimension ecosystem around Melody/Capella modeling workbench for systems engineering (MBSE) and engineering architectures (system, software, hardware).

8.1.2.3. Occiware

- Coordinator: Open Wide
- Open Wide, ActiveEon SA, CSRT - Cloud Systèmes Réseaux et Télécoms, Institut Mines-Télécom/Télécom SudParis, Inria, Linagora, Obeo, OW2 Consortium, Pôle Numérique, Université Joseph Fourier,
- Dates: 2014-2017
- Abstract: The Occiware project aims to establish a formal and equipped framework for the management of all cloud resource based on the OCCI standard.

8.1.3. DGA

8.1.3.1. MOTIV

- Coordinator: InPixal
- Other partners: Bertin, DGA, Inria
- Dates: 2012-2014
- Abstract: This project investigates innovative software test generation and management solutions to handle the very high degrees of variability in video processing algorithmic chains. The objective is to provide systematic criteria to qualify the testing activity when developing video processing software and to tailor these criteria to the variability dimensions that emerge in the context of visible images.

8.1.3.2. FPML (CYBERDEFENSE)

- Coordinator: DGA
- Partners: DGA MI, Inria
- Dates: 2014-2016
- Abstract: in the context of this project, DGA-MI and the Inria team DiverSE explore the existing approaches to ease the development of formal specifications of domain-Specific Languages (DSLs) dedicated to paquet filtering, while guaranteeing expressiveness, precision and safety. In the long term, this work is part of the trend to provide to DGA-MI and its partners a tooling to design and develop formal DSLs which ease the use while ensuring a high level of reasoning.

8.1.4. Cominlabs

8.1.4.1. PROFILE

- Coordinator: Université de Rennes 1
- Partners: Inria, Université de Rennes 2
- Dates: 2016-2019

- Abstract: The PROFILE project brings together experts from law, computer science and sociology to address the challenges raised by online profiling, following a multidisciplinary approach. More precisely, the project will pursue two complementary and mutually informed lines of research: (i) Investigate, design, and introduce a new right of opposition into the legal framework of data protection to better regulate profiling and to modify the behavior of commercial companies towards being more respectful of the privacy of their users; (ii)S Provide users with the technical means they need to detect stealthy profiling techniques as well as to control the extent of the digital traces they routinely produce. As a case study, we focus on browser fingerprinting, a new profiling technique for targeted advertisement. The project will develop a generic framework to reason on the data collected by profiling algorithms, to uncover their inner working, and make them more accountable to users. PROFILE will also propose an innovative protection to mitigate browser fingerprinting, based on the collaborative reconfiguration of browsers.

8.2. European Initiatives

8.2.1. FP7 & H2020 Projects

8.2.1.1. FP7 FET STREP DIVERSIFY

- Coordinator: Inria (DIVERSE)
- Partners: SINTEF, Université de Rennes 1, Trinity College Dublin, Inria (DiverSE, SPIRALS)
- Dates: 2013-2016
- Abstract: DIVERSIFY explores diversity as the foundation for a novel software design principle and increased adaptive capacities in CASs. Higher levels of diversity in the system provide a pool of software solutions that can eventually be used to adapt to unforeseen situations at design time. The scientific development of DIVERSIFY is based on a strong analogy with ecological systems, biodiversity, and evolutionary ecology. DIVERSIFY brings together researchers from the domains of software-intensive distributed systems and ecology in order to translate ecological concepts and processes into software design principles.

8.2.1.2. FP7 STREP HEADS

- Coordinator: SINTEF
- Other partners: Inria, Software AG, ATC, Tellu, eZmonitoring
- Dates: 2013-2016
- Abstract: The idea of the HEADS project is to leverage model-driven software engineering and generative programming techniques to provide a new integrated software engineering approach which allow advanced exploitation the full range of diversity and specificity of the future computing continuum. The goal is to empower the software and services industry to better take advantage of the opportunities of the future computing continuum and to effectively provide new innovative services that are seamlessly integrated to the physical world making them more pervasive, more robust, more reactive and closer (physically, socially, emotionally, etc.) to their users. We denote such services HD-services. HD-services (Heterogeneous and Distributed services) characterize the class of services or applications within the Future Internet whose logic and value emerges from a set of communicating software components distributed on a heterogeneous computing continuum from clouds to mobile devices, sensors and/or smart-objects.

8.2.1.3. H2020 ICT-10-2016 STAMP

- Coordinator: Inria (DIVERSE)
- Other partners: ATOS, ActiveEon, OW2, TellU, Engineering, XWiki, TU Delft, SINTEF
- Dates: 2016-2019

- Abstract: Leveraging advanced research in automatic test generation, STAMP aims at pushing automation in DevOps one step further through innovative methods of test amplification. It will reuse existing assets (test cases, API descriptions, dependency models), in order to generate more test cases and test configurations each time the application is updated. Acting at all steps of development cycle, STAMP techniques aim at reducing the number and cost of regression bugs at unit level, configuration level and production stage.

  STAMP will raise confidence and foster adoption of DevOps by the European IT industry. The project gathers 3 academic partners with strong software testing expertise, 5 software companies (in: e-Health, Content Management, Smart Cities and Public Administration), and an open source consortium. This industry-near research addresses concrete, business-oriented objectives. All solutions are open source and developed as microservices to facilitate exploitation, with a target at TRL 6.

8.2.2. Collaborations in European Programs, Except FP7 & H2020

8.2.2.1. ICT COST Action MPM4CPS (IC1404)

- Chair of the Action: Prof Hans Vangheluwe (BE)
- Dates: 2014-2018
- Abstract: Truly complex, designed systems, known as Cyber Physical Systems (CPS), are emerging that integrate physical, software, and network aspects. To date, no unifying theory nor systematic design methods, techniques and tools exist for such systems. Individual (mechanical, electrical, network or software) engineering disciplines only offer partial solutions. Multi-paradigm Modelling (MPM) proposes to model every part and aspect of a system explicitly, at the most appropriate level(s) of abstraction, using the most appropriate modelling formalism(s). Modelling languages' engineering, including model transformation, and the study of their semantics, are used to realize MPM. MPM is seen as an effective answer to the challenges of designing CPS. This COST Action promotes the sharing of foundations, techniques and tools, and provide educational resources, to both academia and industry. This is achieved by bringing together and disseminating knowledge and experiments on CPS problems and MPM solutions. Benoit Combemale is a member of the management committee.

8.2.3. Collaborations with Major European Organizations

SINTEF, ICT (Norway): Model-driven systems development for the construction of distributed, heterogeneous applications. We collaborate since 2008 and are currently in two FP7 projects together.

Université du Luxembourg, (Luxembourg): Models runtime for dynamic adaptation and multi-objective elasticity in cloud management; model-driven development.

Open University (UK): models runtime for the Internet of Things.

8.3. International Initiatives

- Université de Montréal (Canada)
- McGill University (Canada)
- University of Alabama (USA)
- TU Wien (Austria)
- Michigan State University (USA)
- Aachen University (Germany)

8.3.1. Participation in Other International Programs

The GEMOC studio has been sustained through the creation of a Research Consortium at the Eclipse Foundation.

8.3.2. International initiative GEMOC

The GEMOC initiative (cf. http://www.gemoc.org) is an open and international initiative launched in 2013 that coordinate research partners worldwide to develop breakthrough software language engineering (SLE) approaches that support global software engineering through the use of multiple domain-specific languages. GEMOC members aim to provide effective SLE solutions to problems associated with the design and implementation of collaborative, interoperable and composable modeling languages.

The GEMOC initiative aims to provide a framework that facilitates collaborative work on the challenges of using of multiple domain-specific languages in software development projects. The framework consists of mechanisms for coordinating the work of members, and for disseminating research results and other related information on GEMOC activities. The framework also provides the required infrastructure for sharing artifacts produced by members, including publications, case studies, and tools.

The governance of the GEMOC initiative is ensured by the Advisory Board. The role of the Advisory Board is to coordinate the GEMOC work and to ensure proper dissemination of work products and information about GEMOC events (e.g., meetings, workshops).

Benoit Combemale is the co-founder and currently acts as principal coordinator of the GEMOC initiative. Benoit Combemale and Jean-Marc Jézéquel are part of the Advisory Board, and 9 DIVERSE members are part of the GEMOC initiative.

8.4. International Research Visitors

8.4.1. Visits of International Scientists

Yves Le Traon, Professor at the University of Luxembourg, visited the team in June and July 2016.

Tanja Mayerhofer, Junior Researcher at the TU Wien, visited the team in September 2016.

Bernhard Rumpe, Professor at Aachen University, visited the team in May 2016.

8.4.1.1. Internships

Vikas Mishra, Master intership at the Birla Institute of Technology & Science, visited the team from June to August 2016.

Alexandre Nuttinck, Axel Halin, Master interships at the University of Namur, visited the team from September 2016 to January 2017.

8.4.2. Visits to International Teams

Manuel Leduc visited CWI for 3 weeks in December 2016

Benoit Baudry visited Professor Stephanie Forrest at the University of New Mexico for one month in April 2016.

Benoit Combemale visited Professor Jorg Kienzle at McGill University for 3 weeks in June 2016; and visited Professor Bernhard Rumpe at Aachen University in April 2016.

8.4.2.1. Research Stays Abroad

Marcelino Rodriguez-Cancio visited Vanderbildt University from November 2016 to May 2017.

9. Dissemination

9.1. Promoting Scientific Activities

9.1.1. Scientific Events Organisation

9.1.1.1. General Chair, Scientific Chair

Benoit Baudry and Benoit Combemale have been general co-chairs for MODELS 2016, the major international conference in the area of model-driven software and systems engineering.

9.1.1.2. Member of the Organizing Committees

Mathieu Acher:

- Social media chair, MODELS 2016

Arnaud Blouin:

- Student volunteers co-chair, MODELS 2016
- Social Media Chair, EICS 2016

9.1.2. Scientific Events Selection

9.1.2.1. Chair of Conference Program Committees

Noël Plouzeau was the program committee chair of the Component Based Software Engineering (CBSE) in 2016. Jean-Marc Jézéquel was co-chair of Digital Intelligence 2016, April 2016, Quebec City, Canada.

9.1.2.2. Member of the Conference Program Committees

Mathieu Acher:

- PC member SPLC 2016
- PC member SAC 2016
- PC member VaMoS 2016
- PC member SEAA 2016
- PC member REVE 2016

Olivier Barais:

- PC member SEAA 2016
- PC member 10th IEEE ICOSST-2016
- PC member compas2016
- PC member ASE 2016
- PC member TTC 2016
- PC member SAC 2016

Benoit Baudry:

- PC member ICSE 2016
- PC member ISSTA 2016
- PC member ICST 2016
- PC member ASE 2016

Arnaud Blouin:

- PC member of the foundation track for MODELS 2016
- PC member for IHM 2016
- PC member of the NIER/demo track for VISSOFT 2016
- PC member for the UCDAS workshop at COMPSAC 2016

Benoit Combemale:

- PC member for ECMFA'16
- PC member for ICMT'16
- PC member for the GEMOC'16 workshop at MODELS'16
- PC member for the EXE'16 workshop at MODELS'16
- PC member for the MiSE'16 workshop at ICSE'16
- PC member for the LaMOD'16 workshop at Modularity'16
- PC member for the MoMo'16 workshop at Modularity'16

Johann Bourcier:

- PC member for the SEsCPS'16 workshop at ICSE'16

Jean-Marc Jézéquel:

- Program Board of MODELS 2016
- PC member for SPLC 2016 Industry Track
- PC member for SPLC 2016 Vision Track
- PC member for SEAMS 2016
- PC member for GEMOC 2016

9.1.2.3. Reviewer

Arnaud Blouin was a reviewer for ICST 2016, ASE 2016, ECMFA 2016. Noël Plouzeau was a reviewer for ICSE 2016. Olivier Barais was a reviewer for ICSE 2016, Sosym 2016. Johann Bourcier was a reviewer for ICSE 2016, ASE 2016 and Middleware 2016.

9.1.3. Journal

9.1.3.1. Member of the Editorial Boards

Benoit Baudry:

- SOSYM
- STVR

Benoit Combemale:

- Springer Journal on Software and System Modeling (SoSYM),
- Elsevier Journal on Computer Languages, Systems and Structures (COMLAN),
- Elsevier Journal on Science of Computer Programming (SCP), Advisory Board member of the Software Section.

9.1.3.2. Reviewer - Reviewing Activities

Mathieu Acher was a reviewer for TSE, JSS, IST.

Johann Bourcier was a reviewer for JSS, IEEE Communications Magazine,

Arnaud Blouin: SoSyM

Jean-Marc Jézéquel:

- Associate Editor in Chief of SOSYM
- IEEE Computer
- JSS
- JOT

9.1.4. Invited Talks

Benoit Baudry:

- Running software in uncertain environments. Invited talk at the ACCESS Distinguished Lecture Series at KTH, Stockholm.
- ECSLER: tool support for runtime evolution inside the JVM at the CREST workshop on genetic improvement, University College London.
- Improving software quality and DevOps automation with STAMP at OW2'con 2016.

Benoit Combemale:

- Towards Language-Oriented Modeling. Invited talk at RWTH Aachen University, Germany.
- Dynamic Validation & Verification in Language-Oriented Modeling. Keynote at MoDeVVA 2016.
- On the Globalization of Modeling Languages. Invited talk at CNES.
- A Bit of Software Engineering! Invited talk at Airbus.
- Modeling for Smart Cyber-Physical Systems. Invited talk at the French CPS days.
- Modeling for Sustainability. Invited talk at Lancaster University, United Kingdom.

9.1.5. Leadership within the Scientific Community

Benoit Baudry:

- Steering committee member for the ACM/IEEE MODELS conference

Benoit Combemale:

- Steering committee member for the ACM SLE conference
- Founding member and member of the advisory board of the GEMOC initiative.

9.1.6. Research Administration

Benoit Baudry is the leader of the DiverSE research team; he is in the scientific advisory board of the SVV lab, University of Luxembourg.

Jean-Marc Jézéquel is Director of IRISA (UMR 6074). He is Coordinator of the academic club of the French Cyber-defense Excellence Cluster, and Director of the Rennes Node of EIT Digital.

9.2. Teaching - Supervision - Juries

9.2.1. Teaching

The DIVERSE team bears the bulk of the teaching on Software Engineering at the University of Rennes 1 and at INSA Rennes, for the first year of the Master of Computer Science (Project Management, Object-Oriented Analysis and Design with UML, Design Patterns, Component Architectures and Frameworks, Validation & Verification, Human-Computer Interaction) and for the second year of the MSc in software engineering (Model driven Engineering, Aspect-Oriented Software Development, Software Product Lines, Component Based Software Development, Validation & Verification, *etc.*).

Each of Jean-Marc Jézéquel, Noël Plouzeau, Olivier Barais, Johann Bourcier, Arnaud Blouin, and Mathieu Acher teaches about 200h in these domains, with Benoit Baudry and Benoit Combemale teaching about 50h, for a grand total of about 1300 hours, including several courses at ENSTB, Supelec, and ENSAI Rennes engineering school.

Olivier Barais is deputy director of the electronics and computer science teaching department of the University of Rennes 1. Mathieu Acher is in charge of teaching duties management of this department. Noël Plouzeau is the head of the final year of the Master in Computer Science at the University of Rennes 1. Johann Bourcier is co-manager of the Home-Automation option at the ESIR engineering school in Rennes.

The DIVERSE team also hosts several MSc and summer trainees every year.

9.2.2. Supervision

- PhD: Sana Ben Nasr [17], *Mining and Modeling Variability from Natural Language Documents: Two Case Studies*, Inria, April 2016, B. Baudry and M. Acher
- PhD: Guillaume Becan [19] *Metamodels and Feature Models: Complementary Approaches to Formalize Product Comparison Matrices*, MESR, September 2016, B. Baudry and M. Acher
- PhD: Thomas Degueule , *Composition and interoperability of domain-specific language engineering*, Univ. Rennes I, 02/12/16
- PhD: Francisco Acosta , *Automatic Deployment and Reconfigurations in Internet of Things*, Univ. Rennes I, 12/12/2016
- PhD: David Mendez Acuna , *Variability in Modeling Languages*, Inria, 16/12/2016
- PhD: Jacky Bourgeois , *Automatic Synchronization of Energy Production and Consumption*, Open University, DATE
- PhD in progress: Mohamed Boussaa, *An Architecture for Testing Large-Scale Dynamic Distributed Systems*, 2013, B. Baudry, O. Barais
- PhD in progress: Alejandro Gomez Boix, *Distributed counter-measure against browser fingerprinting*, 2016, B. Baudry, D. Bromberg
- PhD in progress: Nicolas Harrand, *Automatic diversity for code obfuscation*, 2016, B. Baudry
- PhD in progress: Pierre Laperdrix, *Secretless moving target against browser fingerprinting*, B. Baudry, G. Avoine
- PhD in progress: Johan Pelay, *Langage pour une programmation incrémentale de réseau*, 2016, O. Barais, F. Guillemin
- PhD in progress: Quentin Plazar, *Combining Decision Procedures for Constraint Programming and SMT*, 2015, M. Acher, A. Goetib
- PhD in progress: Marcelino Rodriguez Cancio, *Automatic computation diversification*, 2015, B. Baudry and B. Combemale
- PhD in progress: Paul Temple, *Variability Testing of Computer Vision Algorithms*, 2015, M. Acher, J.-M. Jézéquel
- PhD in progress: Kwaku Yeboah-Antwi, *Runtime emergence of software diversity*, 2013, B. Baudry
- PhD in progress: Manuel Leduc, *TITRE*, 2016, O. Barais, B. Combemale
- PhD in progress: Youssou NDiaye, *Modelling and evaluating security in user interfaces*, 2016, N. Aillery, O. Barais, A. Blouin, A. Bouabdallah
- PhD in progress: Jean-Émile Dartois, *TITRE*, 2016, O. Barais
- PhD in progress: Oscar Luis, *Automatic test amplification*, 2016, B. Baudry
- PhD in progress: Alexandre Rio, *TITRE*, 2016, O. Barais, Y. Morel
- PhD: in progress, Kevin Corre, *Modélisation de la confiance dans les services sociaux et conversationnels*, 2014, O. Barais, G. Sunye
- PhD: in progress, Gwendal Le Moulec, *Towards the automatic synthesis of virtual reality applications*, 2015, B. Arnaldi, A. Blouin, V. Gouranton

9.2.3. Juries

9.2.3.1. Mathieu Acher

was in the examination committee of the following PhD thesis:

- Guillaume Becan, September 2016, Univ Rennes I, Supervisor
- Sana Ben Nasr, April 2016, Univ Rennes I, Supervisor
- Jaime Chavarriaga, December 2016, Vrije Universiteit Brussel, Reviewer

9.2.3.2. Arnaud Blouin

was in the examination committee of the following PhD thesis:

- Thomas Degueule, December 2016, Univ Rennes I, Supervisor

9.2.3.3. Olivier Barais

was in the examination committee of the following PhD thesis:

- Thomas Degueule, December 2016, Univ Rennes I, Supervisor
- Maxime Colman, December 2016, Univ Lille I, Reviewer
- Alexandre Garnier, December 2016, EMN - Univ Nantes, Reviewer
- Matias Ezequiel Vara Larsen, April 2016, Univ Nice, Reviewer
- Djamel Khelladi, September 2016, UPMC, Reviewer
- Simon Dupont, April 2016, EMN - Univ Nantes, Reviewer

9.2.3.4. Benoit Baudry

was in the examination committee of the following PhD thesis:

- Guillaume Becan, September 2016, Univ Rennes I, Supervisor
- Sana Ben Nasr, April 2016, Univ Rennes I, Supervisor
- David Mendez, December 2016, Univ Rennes I, Supervisor

9.2.3.5. Olivier Barais

was in the examination committee of the following HDR:

- Sébastien Leriche, November 2016, ENAC, Reviewer

9.2.3.6. Jean-Marc Jézéquel

was in the examination committee of the following HDR:

- Tewfik Ziadi, December 2016, UPMC, Reviewer

9.2.3.7. Benoit Combemale

was in the examination committee of the following PhD thesis:

- Thomas Degueule, December 2016, Univ Rennes I, Supervisor
- Blazo Nastov, november 2016, Univ. Montpellier, Reviewer
- Florent Latombe, June 2016, Univ Toulouse, Examiner
- Andreas Wortmann, June 2016, Aachen University, Germany, Reviewer
- Matias Ezequiel Vara Larsen, April 2016, Univ Nice, Examiner

9.2.3.8. Johann Bourcier

was in the examination committee of the following PhD thesis:

- Francisco Javier Acosta Padilla, December 2016, Univ Rennes I, Supervisor

9.2.3.9. Jean-Marc Jézéquel

was in the examination committee of the following PhD thesis:

- Jabier Martinez, October 2016, Univ. Luxemburg, Examiner

9.3. Popularization

- IT industry forums. Our recent work on diversity against browser fingerprinting and the associated web site [0] have been demonstrated in the a major cybersecurity forum (FIC'15 [0]), the largest IT event organized by the European commission (ICT'15 [0]) and two open source forums (OWF'14 [0] and FOSSA'14 [0]).
- General-audience press. Our work on browser fingerprinting has been featured Le Monde [0], and two full articles appeared in the MISC [0] and the Clubic magazines [0].

[0] https://amiunique.org/
[0] https://fic2015.com/home/
[0] https://ec.europa.eu/digital-single-market/en/ict2015
[0] http://www.openworldforum.paris/
[0] https://fossa.inria.fr/
[0] http://abonnes.lemonde.fr/sciences/article/2014/11/24/le-fossa-bazar-techno-participatif_4528568_1650684.html
[0] http://connect.ed-diamond.com/MISC/MISC-081/Le-fingerprinting-une-nouvelle-technique-de-tracage

10. Bibliography

Major publications by the team in recent years

[1] B. BAUDRY, M. MONPERRUS.*The Multiple Facets of Software Diversity: Recent Developments in Year 2000 and Beyond*, in "ACM Computing Surveys", 2015, vol. 48, n^o^ 1, p. 16:1–16:26, https://hal.inria.fr/hal-01182103.

[2] A. BLOUIN, N. MOHA, B. BAUDRY, H. SAHRAOUI, J.-M. JÉZÉQUEL.*Assessing the Use of Slicing-based Visualizing Techniques on the Understanding of Large Metamodels*, in "Information and Software Technology", 2015, vol. 62, p. 124 - 142 [*DOI :* 10.1016/J.INFSOF.2015.02.007], https://hal.inria.fr/hal-01120558.

[3] M. BOUSSAA, O. BARAIS, B. BAUDRY, G. SUNYÉ.*NOTICE: A Framework for Non-functional Testing of Compilers*, in "Proc. of the Int. Conf. on Software Quality, Reliability & Security (QRS)", August 2016, https://hal.archives-ouvertes.fr/hal-01344835.

[4] G. BÉCAN, M. ACHER, B. BAUDRY, S. BEN NASR.*Breathing Ontological Knowledge Into Feature Model Synthesis: An Empirical Study*, in "Empirical Software Engineering", 2015, vol. 21, n^o^ 4, p. 1794–1841 [*DOI :* 10.1007/S10664-014-9357-1], https://hal.inria.fr/hal-01096969.

[5] G. BÉCAN, N. SANNIER, M. ACHER, O. BARAIS, A. BLOUIN, B. BAUDRY.*Automating the Formalization of Product Comparison Matrices*, in "Proc. of the Int. Conf. on Automated Software Engineering (ASE)", September 2014 [*DOI :* 10.1145/2642937.2643000], https://hal.inria.fr/hal-01058440.

[6] B. COMBEMALE, J. DEANTONI, B. BAUDRY, R. B. FRANCE, J.-M. JÉZÉQUEL, J. GRAY.*Globalizing Modeling Languages*, in "IEEE Computer", June 2014, p. 10-13, https://hal.inria.fr/hal-00994551.

[7] B. COMBEMALE, J. DEANTONI, M. E. VARA LARSEN, F. MALLET, O. BARAIS, B. BAUDRY, R. FRANCE.*Reifying Concurrency for Executable Metamodeling*, in "Proc. of the Int. Conf. on Software Language Engineering", October 2013, p. 365-384 [*DOI :* 10.1007/978-3-319-02654-1_20], https://hal.inria.fr/hal-00850770.

[8] J.-M. DAVRIL, E. DELFOSSE, N. HARIRI, M. ACHER, J. CLELANG-HUANG, P. HEYMANS.*Feature Model Extraction from Large Collections of Informal Product Descriptions*, in "Proc. of the Europ. Software Engineering Conf. and the ACM SIGSOFT Symp. on the Foundations of Software Engineering (ESEC/FSE)", September 2013, p. 290-300 [*DOI :* 10.1145/2491411.2491455], https://hal.inria.fr/hal-00859475.

[9] T. DEGUEULE, B. COMBEMALE, A. BLOUIN, O. BARAIS, J.-M. JÉZÉQUEL.*Melange: A Meta-language for Modular and Reusable Development of DSLs*, in "Proc. of the Int. Conf. on Software Language Engineering (SLE)", October 2015, https://hal.inria.fr/hal-01197038.

[10] J. A. GALINDO DUARTE, M. ALFEREZ, M. ACHER, B. BAUDRY, D. BENAVIDES.*A Variability-Based Testing Approach for Synthesizing Video Sequences*, in "Proc. of the Int. Symp. on Software Testing and Analysis (ISSTA)", July 2014, https://hal.inria.fr/hal-01003148.

[11] I. GONZALEZ-HERRERA, J. BOURCIER, E. DAUBERT, W. RUDAMETKIN, O. BARAIS, F. FOUQUET, J.-M. JÉZÉQUEL, B. BAUDRY.*ScapeGoat: Spotting abnormal resource usage in component-based reconfigurable*

[0] http://pro.clubic.com/webmarketing/publicite-en-ligne/actualite-742853-fingerprinting-cookies.html

software systems, in "Journal of Systems and Software", 2016 [*DOI :* 10.1016/J.JSS.2016.02.027], https://hal.inria.fr/hal-01354999.

[12] J.-M. JÉZÉQUEL, B. COMBEMALE, O. BARAIS, M. MONPERRUS, F. FOUQUET.*Mashup of Meta-Languages and its Implementation in the Kermeta Language Workbench*, in "Software and Systems Modeling", 2015, vol. 14, n^{o} 2, p. 905-920, https://hal.inria.fr/hal-00829839.

[13] P. LAPERDRIX, W. RUDAMETKIN, B. BAUDRY.*Beauty and the Beast: Diverting modern web browsers to build unique browser fingerprints*, in "Proc. of the Symp. on Security and Privacy (S&P)", May 2016, https://hal.inria.fr/hal-01285470.

[14] M. RODRIGUEZ-CANCIO, B. COMBEMALE, B. BAUDRY.*Automatic Microbenchmark Generation to Prevent Dead Code Elimination and Constant Folding*, in "Proc. of the Int. Conf. on Automated Software Engineering (ASE)", September 2016, https://hal.inria.fr/hal-01343818.

[15] M. TRICOIRE, O. BARAIS, M. LEDUC, J. BOURCIER, F. FOUQUET, G. NAIN, M. LUDOVIC, G. SUNYÉ, B. MORIN.*KevoreeJS: Enabling Dynamic Software Reconfigurations in the Browser*, in "Proc. of WICSA and CompArch", April 2016 [*DOI :* 10.1109/CBSE.2016.20], https://hal.inria.fr/hal-01354997.

Publications of the year

Doctoral Dissertations and Habilitation Theses

[16] F. J. ACOSTA PADILLA.*Self-adaptation for Internet of Things applications*, Université de Rennes 1, France, December 2016, https://hal.inria.fr/tel-01426219.

[17] S. BEN NASR.*Mining and Modeling Variability from Natural Language Documents: Two Case Studies*, Université Rennes 1, April 2016, https://hal.inria.fr/tel-01388392.

[18] J. BOURGEOIS.*Interactive Demand-Shifting in the Context of Domestic Micro-Generation*, The Open University ; Université de Rennes 1 [UR1], June 2016, https://hal.inria.fr/tel-01385022.

[19] G. BÉCAN.*Metamodels and Feature Models: Complementary Approaches to Formalize Product Comparison Matrices*, Université Rennes 1, September 2016, https://hal.inria.fr/tel-01416129.

[20] T. DEGUEULE.*Composition and Interoperability for External Domain-Specific Language Engineering*, Université de Rennes 1 [UR1], December 2016, https://hal.inria.fr/tel-01427009.

[21] D. MÉNDEZ-ACUÑA.*Leveraging Software Product Lines Engineering in the Construction of Domain Specific Languages*, Université de Rennes 1, France, December 2016, https://hal.archives-ouvertes.fr/tel-01427187.

Articles in International Peer-Reviewed Journal

[22] S. BEN NASR, G. BÉCAN, M. ACHER, J. F. F. BOSCO, N. SANNIER, B. BAUDRY, J.-M. DAVRIL.*Automated Extraction of Product Comparison Matrices From Informal Product Descriptions*, in "Journal of Systems and Software", 2017, vol. 124, p. 82 - 103 [*DOI :* 10.1016/J.JSS.2016.11.018], https://hal.inria.fr/hal-01427218.

[23] J. F. F. BOSCO, M. ACHER, O. BARAIS.*Software Unbundling: Challenges and Perspectives*, in "Transactions on Modularity and Composition", May 2016, https://hal.inria.fr/hal-01427560.

[24] T. DEGUEULE, B. COMBEMALE, A. BLOUIN, O. BARAIS, J.-M. JÉZÉQUEL. *Safe Model Polymorphism for Flexible Modeling*, in "Computer Languages, Systems and Structures", September 2016, https://hal.inria.fr/hal-01367305.

[25] I. GONZALEZ-HERRERA, J. BOURCIER, E. DAUBERT, W. RUDAMETKIN, O. BARAIS, F. FOUQUET, J.-M. JÉZÉQUEL, B. BAUDRY. *ScapeGoat: Spotting abnormal resource usage in component-based reconfigurable software systems*, in "Journal of Systems and Software", 2016 [*DOI :* 10.1016/J.JSS.2016.02.027], https://hal.inria.fr/hal-01354999.

[26] A. HERVIEU, D. MARIJAN, A. GOTLIEB, B. BAUDRY. *Optimal Minimisation of Pairwise-covering Test Configurations Using Constraint Programming*, in "Information and Software Technology", March 2016, vol. 71, p. 129-146, https://hal.inria.fr/hal-01352831.

[27] M. KEZADRI, M. PANTEL, X. THIRIOUX, B. COMBEMALE. *Correct-by-construction model driven engineering composition operators*, in "Formal Aspects of Computing", January 2016, vol. 28, n^o 3 [*DOI :* 10.1007/S00165-016-0354-6], https://hal.inria.fr/hal-01319576.

[28] G. KHANDU NARWANE, J. A. GALINDO DUARTE, S. NARAYANAN KRISHNA, D. BENAVIDES, J.-V. MILLO, S. RAMESH. *Traceability analyses between features and assets in software product lines*, in "Entropy", July 2016, https://hal.inria.fr/hal-01342351.

[29] D. MÉNDEZ-ACUÑA, J. A. GALINDO DUARTE, T. DEGUEULE, B. COMBEMALE, B. BAUDRY. *Leveraging Software Product Lines Engineering in the Development of External DSLs: A Systematic Literature Review*, in "Computer Languages, Systems and Structures", September 2016 [*DOI :* 10.1016/J.CL.2016.09.004], https://hal.archives-ouvertes.fr/hal-01372702.

[30] J. XUAN, B. CORNU, M. MARTINEZ, B. BAUDRY, L. SEINTURIER, M. MONPERRUS. *B-Refactoring: Automatic Test Code Refactoring to Improve Dynamic Analysis*, in "Information and Software Technology", 2016, vol. 76, p. 65-80 [*DOI :* 10.1016/J.INFSOF.2016.04.016], https://hal.archives-ouvertes.fr/hal-01309004.

[31] K. YEBOAH-ANTWI, B. BAUDRY. *Online Genetic Improvement on the java virtual machine with ECSELR*, in "Genetic Programming and Evolvable Machines", October 2016, p. 1-27 [*DOI :* 10.1007/S10710-016-9278-4], https://hal.inria.fr/hal-01382964.

Invited Conferences

[32] G. PERROUIN, M. ACHER, J.-M. DAVRIL, A. LEGAY, P. HEYMANS. *A Complexity Tale: Web Configurators*, in "VACE 2016 - 1st International Workshop on Variability and Complexity in Software Design Pages (co-located with ICSE'16)", Austin, United States, May 2016, p. 28 - 31 [*DOI :* 10.1145/2897045.2897051], https://hal.inria.fr/hal-01427165.

International Conferences with Proceedings

[33] A. ALLEN, C. ARAGON, C. BECKER, J. C. CARVER, A. CHIS, B. COMBEMALE, M. CROUCHER, K. CROWSTON, D. GARIJO, A. GEHANI, C. GOBLE, R. HAINES, R. HIRSCHFELD, J. HOWISON, K. HUFF, C. JAY, D. KATZ, C. KIRCHNER, K. KUKSENOK, R. LÄMMEL, O. NIERSTRASZ, M. TURK, R. V. VAN NIEUWPOORT, M. VAUGHN, J. VINJU. *Lightning Talk:"I solemnly pledge" A Manifesto for Personal Responsibility in the Engineering of Academic Software*, in "Fourth Workshop on Sustainable Software for Science: Practice and Experiences (WSSSPE4)", Manchester, United Kingdom, G. ALLEN, J. CARVER, M. HEROUX, L. J. HWANG, D. S. KATZ, K. E. NIEMEYER, M. PARASHAR, C. C. VENTERS, S.-C. T.

CHOI, T. CRICK, M. R. CRUSOE, S. GESING, R. HAINES (editors), Proceedings of the Fourth Workshop on Sustainable Software for Science: Practice and Experiences (WSSSPE4), CEUR, 2016, vol. 1686, https://hal.inria.fr/hal-01367344.

[34] O. BARAIS, J. BOURCIER, Y.-D. BROMBERG, C. DION.*Towards microservices architecture to transcode videos in the large at low costs*, in "TEMU 2016 - International Conference on Telecommunications and Multimedia", Heraklion, Greece, July 2016, p. 1 - 6 [*DOI :* 10.1109/TEMU.2016.7551918], https://hal.inria.fr/hal-01427277.

[35] M. BOUSSAA, O. BARAIS, B. BAUDRY, G. SUNYÉ.*Automatic Non-functional Testing of Code Generators Families*, in "15th International Conference on Generative Programming: Concepts & Experiences (GPCE 2016)", Amsterdam, Netherlands, Proceedings of the 2016 ACM SIGPLAN International Conference on Generative Programming: Concepts and Experiences, October 2016, 11, https://hal.inria.fr/hal-01356849.

[36] M. BOUSSAA, O. BARAIS, B. BAUDRY, G. SUNYÉ.*NOTICE: A Framework for Non-functional Testing of Compilers*, in "2016 IEEE International Conference on Software Quality, Reliability & Security (QRS 2016)", Vienna, Austria, IEEE Computer Society Conference Publishing Services (CPS), August 2016, https://hal.archives-ouvertes.fr/hal-01344835.

[37] E. BOUSSE, T. DEGUEULE, D. VOJTISEK, T. MAYERHOFER, J. DEANTONI, B. COMBEMALE.*Execution Framework of the GEMOC Studio (Tool Demo)*, in "Proceedings of the 2016 ACM SIGPLAN International Conference on Software Language Engineering", Amsterdam, Netherlands, SLE 2016, October 2016, 8, https://hal.inria.fr/hal-01355391.

[38] B. COMBEMALE, C. BRUN, J. CHAMPEAU, X. CRÉGUT, J. DEANTONI, J. LE NOIR.*A Tool-Supported Approach for Concurrent Execution of Heterogeneous Models*, in "8th European Congress on Embedded Real Time Software and Systems (ERTS 2016)", Toulouse, France, 2016, https://hal.inria.fr/hal-01258358.

[39] B. COMBEMALE, B. H. CHENG, A. MOREIRA, J.-M. BRUEL, J. GRAY.*Modeling for Sustainability*, in "Modeling in Software Engineering 2016 (MiSE'16)", Austin, United States, Modeling in Software Engineering 2016 (MiSE'16), ACM, 2016, https://hal.inria.fr/hal-01185800.

[40] J. A. GALINDO DUARTE, M. ACHER, J. M. TIRADO, C. VIDAL, B. BAUDRY, D. BENAVIDES.*Exploiting the Enumeration of All Feature Model Configurations: A New Perspective with Distributed Computing*, in "Software Product Line Conference", Beijing, China, September 2016, https://hal.archives-ouvertes.fr/hal-01334851.

[41] I. GONZALEZ-HERRERA, J. BOURCIER, W. RUDAMETKIN, O. BARAIS, F. FOUQUET.*Squirrel: Architecture Driven Resource Management*, in "SAC - 31st Annual ACM Symposium on Applied Computing", Pisa, Italy, April 2016 [*DOI :* 10.1145/0000000.0000000], https://hal.inria.fr/hal-01355000.

[42] A. JANGDA, M. MISHRA, B. BAUDRY.*libmask: Protecting Browser JIT Engines from the Devil in the Constants*, in "Annual Conference on Privacy, Security and Trust", Auckland, New Zealand, December 2016, https://hal.inria.fr/hal-01382971.

[43] J. KIENZLE, G. MUSSBACHER, O. ALAM, M. SCHÖTTLE, N. BELLOIR, P. COLLET, B. COMBEMALE, J. DEANTONI, J. KLEIN, B. RUMPE.*VCU: The Three Dimensions of Reuse*, in "The 15th International Conference on Software Reuse (ICSR-15)", Limassol, Cyprus, The 15th International Conference on Software Reuse, 2016, https://hal.inria.fr/hal-01287720.

[44] J. H. KIM, A. LEGAY, L.-M. TRAONOUEZ, M. ACHER, S. KANG.*A Formal Modeling and Analysis Framework for Software Product Line of Preemptive Real-Time Systems*, in "Symposium on Applied Computing", Pise, Italy, Proceedings of the 31st Annual ACM Symposium on Applied Computing, ACM, April 2016, p. 1562 - 1565 [*DOI :* 10.1145/2851613.2851977], https://hal.archives-ouvertes.fr/hal-01241673.

[45] P. LAPERDRIX, W. RUDAMETKIN, B. BAUDRY.*Beauty and the Beast: Diverting modern web browsers to build unique browser fingerprints*, in "37th IEEE Symposium on Security and Privacy (S&P 2016)", San Jose, United States, May 2016, https://hal.inria.fr/hal-01285470.

[46] G. LE GUERNIC, B. COMBEMALE, J. A. GALINDO DUARTE.*Industrial Experience Report on the Formal Specification of a Packet Filtering Language Using the K Framework*, in "3rd Workshop on Formal Integrated Development Environment", Limassol, Cyprus, Catherine Dubois and Dominique Mery and Paolo Masci, November 2016, https://hal.inria.fr/hal-01401849.

[47] J. LE NOIR, S. MADELÉNAT, C. LABREUCHE, O. CONSTANT, G. GAILLIARD, M. ACHER, O. BARAIS.*A Decision-making Process for Exploring Architectural Variants in Systems Engineering*, in "Software Product Lines Conference (SPLC)", Beijing, China, September 2016 [*DOI :* 10.1145/1235], https://hal.inria.fr/hal-01374140.

[48] V. LELLI, A. BLOUIN, B. BAUDRY, F. COULON, O. BEAUDOUX.*Automatic Detection of GUI Design Smells: The Case of Blob Listener*, in "Proceedings of the 8th ACM SIGCHI Symposium on Engineering Interactive Computing Systems (EICS'16)", Brussels, Belgium, June 2016, p. 263-274 [*DOI :* 10.1145/2933242.2933260], https://hal.inria.fr/hal-01308625.

[49] B. MORIN, F. FLEUREY, K. E. HUSA, O. BARAIS.*A Generative Middleware for Heterogeneous and Distributed Services*, in "19th International ACM Sigsoft Symposium on Component-Based Software Engineering (CBSE 2016)", Venise, Italy, 19th International ACM Sigsoft Symposium on Component-Based Software Engineering, April 2016 [*DOI :* 10.1109/CBSE.2016.12], https://hal.archives-ouvertes.fr/hal-01356104.

[50] D. MÉNDEZ-ACUÑA, J. A. GALINDO DUARTE, B. COMBEMALE, A. BLOUIN, B. BAUDRY, G. LE GUERNIC.*Reverse-engineering reusable language modules from legacy domain-specific languages*, in "International Conference on Software Reuse", Limassol, Cyprus, Proceedings of the International Conference on Software Reuse, June 2016, https://hal.archives-ouvertes.fr/hal-01284816.

[51] D. MÉNDEZ-ACUÑA, J. A. GALINDO DUARTE, B. COMBEMALE, A. BLOUIN, B. BAUDRY.*Puzzle: A tool for analyzing and extracting specification clones in DSLs*, in "ICSR 2016 the 15th International Conference on Software Reuse", Limassol, Cyprus, Proceedings of the 15th International Conference on Software Reuse, June 2016, https://hal.archives-ouvertes.fr/hal-01284822.

[52] E. OUTIN, J.-E. DARTOIS, O. BARAIS, J.-L. PAZAT.*Seeking for the Optimal Energy Modelisation Accuracy to Allow Efficient Datacenter Optimizations*, in "16th International Symposium on Cluster, Cloud and Grid Computing", Cartagena, Italy, IEEE/ACM 16th International Symposium on Cluster, Cloud and Grid Computing, May 2016 [*DOI :* 10.1109/CCGRID.2016.67], https://hal.archives-ouvertes.fr/hal-01356099.

[53] G. PERROUIN, M. AMRANI, M. ACHER, B. COMBEMALE, A. LEGAY, P.-Y. SCHOBBENS.*Featured model types: Towards Systematic Reuse in Modelling Language Engineering*, in "MiSE '16 - 8th International Workshop on Modeling in Software Engineering", New york, United States, ACM, May 2016, p. 1 - 7 [*DOI :* 10.1145/2896982.2896987], https://hal.inria.fr/hal-01406507.

[54] M. RODRIGUEZ-CANCIO, B. COMBEMALE, B. BAUDRY.*Automatic Microbenchmark Generation to Prevent Dead Code Elimination and Constant Folding*, in "31st IEEE/ACM International Conference on Automated Software Engineering (ASE 2016)", Singapore, Singapore, ASE 2016:Proceedings of the 31st IEEE/ACM International Conference on Automated Software Engineering, September 2016, https://hal.inria.fr/hal-01343818.

[55] P. TEMPLE, J. A. GALINDO DUARTE, M. ACHER, J.-M. JÉZÉQUEL.*Using Machine Learning to Infer Constraints for Product Lines*, in "Software Product Line Conference (SPLC)", Beijing, China, September 2016 [*DOI :* 10.1145/2934466.2934472], https://hal.inria.fr/hal-01323446.

[56] M. TRICOIRE, O. BARAIS, M. LEDUC, J. BOURCIER, F. FOUQUET, G. NAIN, M. LUDOVIC, G. SUNYÉ, B. MORIN.*KevoreeJS: Enabling Dynamic Software Reconfigurations in the Browser*, in "WICSA and CompArch 2016", Venice, Italy, ACM Sigsoft, April 2016 [*DOI :* 10.1109/CBSE.2016.20], https://hal.inria.fr/hal-01354997.

[57] M. VINUEZA, J. L. RODAS, J. A. GALINDO DUARTE, D. BENAVIDES.*El uso de modelos de características con atributos para pruebas en sistemas de alta variabilidad: primeros pasos*, in "CEDI 2016", Salamanca, Spain, September 2016, https://hal.inria.fr/hal-01342354.

National Conferences with Proceeding

[58] J. L. RODAS, J. OLIVARES, J. A. GALINDO DUARTE, D. BENAVIDES.*Hacia el uso de sistemas de recomendación en sistemas de alta variabilidad*, in "CEDI 2016", Salamanca, Spain, September 2016, https://hal.inria.fr/hal-01342353.

Scientific Books (or Scientific Book chapters)

[59] B. COMBEMALE, R. FRANCE, J.-M. JÉZÉQUEL, B. RUMPE, J. R. STEEL, D. VOJTISEK.*Engineering Modeling Languages: Turning Domain Knowledge into Tools*, Chapman and Hall/CRC, November 2016, 398, https://hal.inria.fr/hal-01355374.

[60] T. DEGUEULE, B. COMBEMALE, J.-M. JÉZÉQUEL.*On Language Interfaces*, in "PAUSE: Present And Ulterior Software Engineering", B. MEYER, M. MAZZARA (editors), Springer, February 2017, https://hal.inria.fr/hal-01424909.

Research Reports

[61] M. ACHER, F. ESNAULT.*Large-scale Analysis of Chess Games with Chess Engines: A Preliminary Report*, Inria Rennes Bretagne Atlantique, April 2016, n^{o} RT-0479, https://hal.inria.fr/hal-01307091.

[62] B. DANGLOT, P. PREUX, B. BAUDRY, M. MONPERRUS.*Correctness Attraction: A Study of Stability of Software Behavior Under Runtime Perturbation*, HAL, 2016, n^{o} hal-01378523, https://hal.archives-ouvertes.fr/hal-01378523.

[63] G. LE GUERNIC, J. A. GALINDO DUARTE.*Experience Report on the Formal Specification of a Packet Filtering Language Using the K Framework*, Inria Rennes Bretagne Atlantique, October 2016, n^{o} RR-8967, 41, https://hal.inria.fr/hal-01385541.

Other Publications

[64] T. DEGUEULE.*Interoperability and Composition of DSLs with Melange*, June 2016, working paper or preprint, https://hal.inria.fr/hal-01336940.

[65] M. RODRIGUEZ-CANCIO, B. COMBEMALE, B. BAUDRY.*Approximate Loop Unrolling*, November 2016, working paper or preprint, https://hal.inria.fr/hal-01401828.

References in notes

[66] A. ARCURI, L. C. BRIAND.*A practical guide for using statistical tests to assess randomized algorithms in software engineering*, in "ICSE", 2011, p. 1-10.

[67] A. AVIZIENIS.*The N-version approach to fault-tolerant software*, in "Software Engineering, IEEE Transactions on", 1985, n^{o} 12, p. 1491–1501.

[68] F. BACHMANN, L. BASS.*Managing variability in software architectures*, in "SIGSOFT Softw. Eng. Notes", 2001, vol. 26, n^{o} 3, p. 126–132.

[69] F. BALARIN, Y. WATANABE, H. HSIEH, L. LAVAGNO, C. PASSERONE, A. SANGIOVANNI-VINCENTELLI.*Metropolis: An integrated electronic system design environment*, in "Computer", 2003, vol. 36, n^{o} 4, p. 45–52.

[70] E. BANIASSAD, S. CLARKE.*Theme: an approach for aspect-oriented analysis and design*, in "26th International Conference on Software Engineering (ICSE)", 2004, p. 158-167.

[71] E. G. BARRANTES, D. H. ACKLEY, S. FORREST, D. STEFANOVIĆ.*Randomized instruction set emulation*, in "ACM Transactions on Information and System Security (TISSEC)", 2005, vol. 8, n^{o} 1, p. 3–40.

[72] D. BATORY, R. E. LOPEZ-HERREJON, J.-P. MARTIN.*Generating Product-Lines of Product-Families*, in "ASE '02: Automated software engineering", IEEE, 2002, p. 81–92.

[73] S. BECKER, H. KOZIOLEK, R. REUSSNER.*The Palladio component model for model-driven performance prediction*, in "Journal of Systems and Software", January 2009, vol. 82, n^{o} 1, p. 3–22.

[74] N. BENCOMO.*On the use of software models during software execution*, in "MISE '09: Proceedings of the 2009 ICSE Workshop on Modeling in Software Engineering", IEEE Computer Society, May 2009.

[75] A. BEUGNARD, J.-M. JÉZÉQUEL, N. PLOUZEAU.*Contract Aware Components, 10 years after*, in "WCSI", 2010, p. 1-11.

[76] J. BOSCH.*Design and use of software architectures: adopting and evolving a product-line approach*, ACM Press/Addison-Wesley Publishing Co., New York, NY, USA, 2000.

[77] J. BOSCH, G. FLORIJN, D. GREEFHORST, J. KUUSELA, J. H. OBBINK, K. POHL.*Variability Issues in Software Product Lines*, in "PFE '01: Revised Papers from the 4th International Workshop on Software Product-Family Engineering", London, UK, Springer-Verlag, 2002, p. 13–21.

[78] L. C. BRIAND, E. ARISHOLM, S. COUNSELL, F. HOUDEK, P. THÉVENOD–FOSSE.*Empirical studies of object-oriented artifacts, methods, and processes: state of the art and future directions*, in "Empirical Software Engineering", 1999, vol. 4, n^{o} 4, p. 387–404.

[79] J. T. BUCK, S. HA, E. A. LEE, D. G. MESSERSCHMITT.*Ptolemy: A framework for simulating and prototyping heterogeneous systems*, in "Int. Journal of Computer Simulation", 1994.

[80] T. BURES, P. HNETYNKA, F. PLASIL.*Sofa 2.0: Balancing advanced features in a hierarchical component model*, in "Software Engineering Research, Management and Applications, 2006. Fourth International Conference on", IEEE, 2006, p. 40–48.

[81] B. H. C. CHENG, R. LEMOS, H. GIESE, P. INVERARDI, J. MAGEE, J. ANDERSSON, B. BECKER, N. BENCOMO, Y. BRUN, B. CUKIC, G. MARZO SERUGENDO, S. DUSTDAR, A. FINKELSTEIN, C. GACEK, K. GEIHS, V. GRASSI, G. KARSAI, H. M. KIENLE, J. KRAMER, M. LITOIU, S. MALEK, R. MIRANDOLA, H. A. MÜLLER, S. PARK, M. SHAW, M. TICHY, M. TIVOLI, D. WEYNS, J. WHITTLE. , D. HUTCHISON, T. KANADE, J. KITTLER, J. M. KLEINBERG, F. MATTERN, J. C. MITCHELL, M. NAOR, O. NIERSTRASZ, C. PANDU RANGAN, B. STEFFEN, M. SUDAN, D. TERZOPOULOS, D. TYGAR, M. Y. VARDI, G. WEIKUM, B. H. C. CHENG, R. LEMOS, H. GIESE, P. INVERARDI, J. MAGEE (editors)*Software Engineering for Self-Adaptive Systems: A Research Roadmap* , Betty H. C. Cheng, Rogério de Lemos, Holger Giese, Paola Inverardi, and Jeff Magee, Springer Berlin Heidelberg, Berlin, Heidelberg, 2009, vol. 5525.

[82] J. COPLIEN, D. HOFFMAN, D. WEISS.*Commonality and Variability in Software Engineering*, in "IEEE Software", 1998, vol. 15, n^{o} 6, p. 37–45.

[83] I. CRNKOVIC, S. SENTILLES, A. VULGARAKIS, M. R. CHAUDRON.*A classification framework for software component models*, in "Software Engineering, IEEE Transactions on", 2011, vol. 37, n^{o} 5, p. 593–615.

[84] K. CZARNECKI, U. W. EISENECKER.*Generative programming: methods, tools, and applications*, ACM Press/Addison-Wesley Publishing Co., New York, NY, USA, 2000.

[85] R. DEMILLI, A. J. OFFUTT.*Constraint-based automatic test data generation*, in "Software Engineering, IEEE Transactions on", 1991, vol. 17, n^{o} 9, p. 900–910.

[86] K. DEB, A. PRATAP, S. AGARWAL, T. MEYARIVAN.*A fast and elitist multiobjective genetic algorithm: NSGA-II*, in "Evolutionary Computation, IEEE Transactions on", 2002, vol. 6, n^{o} 2, p. 182–197.

[87] S. FORREST, A. SOMAYAJI, D. H. ACKLEY.*Building diverse computer systems*, in "Operating Systems, 1997., The Sixth Workshop on Hot Topics in", IEEE, 1997, p. 67–72.

[88] R. B. FRANCE, B. RUMPE.*Model-driven Development of Complex Software: A Research Roadmap*, in "Proceedings of the Future of Software Engineering Symposium (FOSE '07)", L. C. BRIAND, A. L. WOLF (editors), IEEE, 2007, p. 37–54.

[89] S. FREY, F. FITTKAU, W. HASSELBRING.*Search-based genetic optimization for deployment and reconfiguration of software in the cloud*, in "Proceedings of the 2013 International Conference on Software Engineering", IEEE Press, 2013, p. 512–521.

[90] G. HALMANS, K. POHL.*Communicating the Variability of a Software-Product Family to Customers*, in "Software and System Modeling", 2003, vol. 2, n^{o} 1, p. 15-36.

[91] C. HARDEBOLLE, F. BOULANGER.*ModHel'X: A component-oriented approach to multi-formalism modeling*, in "Models in Software Engineering", Springer, 2008, p. 247–258.

[92] M. HARMAN, B. F. JONES.*Search-based software engineering*, in "Information and Software Technology", 2001, vol. 43, n^{o} 14, p. 833–839.

[93] H. HEMMATI, L. C. BRIAND, A. ARCURI, S. ALI.*An enhanced test case selection approach for model-based testing: an industrial case study*, in "SIGSOFT FSE", 2010, p. 267-276.

[94] J. HUTCHINSON, J. WHITTLE, M. ROUNCEFIELD, S. KRISTOFFERSEN.*Empirical assessment of MDE in industry*, in "Proceedings of the 33rd International Conference on Software Engineering (ICSE '11)", R. N. TAYLOR, H. GALL, N. MEDVIDOVIC (editors), ACM, 2011, p. 471–480.

[95] J.-M. JÉZÉQUEL.*Model Driven Design and Aspect Weaving*, in "Journal of Software and Systems Modeling (SoSyM)", may 2008, vol. 7, n^{o} 2, p. 209–218, http://www.irisa.fr/triskell/publis/2008/Jezequel08a.pdf.

[96] K. C. KANG, S. G. COHEN, J. A. HESS, W. E. NOVAK, A. S. PETERSON.*Feature-Oriented Domain Analysis (FODA) Feasibility Study*, Carnegie-Mellon University Software Engineering Institute, November 1990.

[97] J. KRAMER, J. MAGEE.*Self-Managed Systems: an Architectural Challenge*, in "Future of Software Engineering", IEEE, 2007, p. 259–268.

[98] K.-K. LAU, P. V. ELIZONDO, Z. WANG.*Exogenous connectors for software components*, in "Component-Based Software Engineering", Springer, 2005, p. 90–106.

[99] P. MCMINN.*Search-based software test data generation: a survey*, in "Software Testing, Verification and Reliability", 2004, vol. 14, n^{o} 2, p. 105–156.

[100] J. MEEKEL, T. B. HORTON, C. MELLONE.*Architecting for Domain Variability*, in "ESPRIT ARES Workshop", 1998, p. 205-213.

[101] A. M. MEMON.*An event-flow model of GUI-based applications for testing*, in "Software Testing, Verification and Reliability", 2007, vol. 17, n^{o} 3, p. 137–157.

[102] B. MORIN, O. BARAIS, J.-M. JÉZÉQUEL, F. FLEUREY, A. SOLBERG.*Models at Runtime to Support Dynamic Adaptation*, in "IEEE Computer", October 2009, p. 46-53, http://www.irisa.fr/triskell/publis/2009/Morin09f.pdf.

[103] P.-A. MULLER, F. FLEUREY, J.-M. JÉZÉQUEL.*Weaving Executability into Object-Oriented Meta-Languages*, in "Proc. of MODELS/UML'2005", Jamaica, LNCS, Springer, 2005.

[104] R. MÉLISSON, P. MERLE, D. ROMERO, R. ROUVOY, L. SEINTURIER.*Reconfigurable run-time support for distributed service component architectures*, in "the IEEE/ACM international conference", New York, New York, USA, ACM Press, 2010, 171.

[105] C. NEBUT, Y. LE TRAON, J.-M. JÉZÉQUEL.*System Testing of Product Families: from Requirements to Test Cases*, Springer Verlag, 2006, p. 447–478, http://www.irisa.fr/triskell/publis/2006/Nebut06b.pdf.

[106] C. NEBUT, S. PICKIN, Y. LE TRAON, J.-M. JÉZÉQUEL.*Automated Requirements-based Generation of Test Cases for Product Families*, in "Proc. of the 18th IEEE International Conference on Automated Software Engineering (ASE'03)", 2003, http://www.irisa.fr/triskell/publis/2003/nebut03b.pdf.

[107] L. M. NORTHROP.*SEI's Software Product Line Tenets*, in "IEEE Softw.", 2002, vol. 19, n° 4, p. 32–40.

[108] L. M. NORTHROP.*A Framework for Software Product Line Practice*, in "Proceedings of the Workshop on Object-Oriented Technology", Springer-Verlag London, UK, 1999, p. 365–376.

[109] I. OBER, S. GRAF, I. OBER.*Validating timed UML models by simulation and verification*, in "International Journal on Software Tools for Technology Transfer", 2006, vol. 8, n° 2, p. 128–145.

[110] D. L. PARNAS.*On the Design and Development of Program Families*, in "IEEE Trans. Softw. Eng.", 1976, vol. 2, n° 1, p. 1–9.

[111] S. PICKIN, C. JARD, T. JÉRON, J.-M. JÉZÉQUEL, Y. LE TRAON.*Test Synthesis from UML Models of Distributed Software*, in "IEEE Transactions on Software Engineering", April 2007, vol. 33, n° 4, p. 252–268.

[112] K. POHL, G. BÖCKLE, F. J. VAN DER LINDEN.*Software Product Line Engineering: Foundations, Principles and Techniques*, Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2005.

[113] B. RANDELL.*System structure for software fault tolerance*, in "Software Engineering, IEEE Transactions on", 1975, n° 2, p. 220–232.

[114] M. RINARD.*Obtaining and reasoning about good enough software*, in "Proceedings of Annual Design Automation Conference (DAC)", 2012, p. 930-935.

[115] J. ROTHENBERG, L. E. WIDMAN, K. A. LOPARO, N. R. NIELSEN.*The Nature of Modeling*, in "in Artificial Intelligence, Simulation and Modeling", John Wiley & Sons, 1989, p. 75–92.

[116] P. RUNESON, M. HÖST.*Guidelines for conducting and reporting case study research in software engineering*, in "Empirical Software Engineering", 2009, vol. 14, n° 2, p. 131–164.

[117] D. SCHMIDT.*Guest Editor's Introduction: Model-Driven Engineering*, in "IEEE Computer", 2006, vol. 39, n° 2, p. 25–31.

[118] F. SHULL, J. SINGER, D. I. SJBERG.*Guide to advanced empirical software engineering*, Springer, 2008.

[119] S. SIDIROGLOU-DOUSKOS, S. MISAILOVIC, H. HOFFMANN, M. RINARD.*Managing performance vs. accuracy trade-offs with loop perforation*, in "Proc. of the Symp. on Foundations of software engineering", New York, NY, USA, ESEC/FSE '11, ACM, 2011, p. 124-134.

[120] J. STEEL, J.-M. JÉZÉQUEL.*On Model Typing*, in "Journal of Software and Systems Modeling (SoSyM)", December 2007, vol. 6, n° 4, p. 401–414, http://www.irisa.fr/triskell/publis/2007/Steel07a.pdf.

[121] C. SZYPERSKI, D. GRUNTZ, S. MURER.*Component software: beyond object-oriented programming*, Addison-Wesley, 2002.

[122] J.-C. TRIGAUX, P. HEYMANS.*Modelling variability requirements in Software Product Lines: a comparative survey*, FUNDP Namur, 2003.

[123] M. UTTING, B. LEGEARD.*Practical model-based testing: a tools approach*, Morgan Kaufmann, 2010.

[124] P. VROMANT, D. WEYNS, S. MALEK, J. ANDERSSON.*On interacting control loops in self-adaptive systems*, in "SEAMS 2011", ACM, 2011, p. 202–207.

[125] C. YILMAZ, M. B. COHEN, A. A. PORTER.*Covering arrays for efficient fault characterization in complex configuration spaces*, in "Software Engineering, IEEE Transactions on", 2006, vol. 32, n^{o} 1, p. 20–34.

[126] Z. A. ZHU, S. MISAILOVIC, J. A. KELNER, M. RINARD.*Randomized accuracy-aware program transformations for efficient approximate computations*, in "Proc. of the Symp. on Principles of Programming Languages (POPL)", 2012, p. 441-454.

[127] T. ZIADI, J.-M. JÉZÉQUEL.*Product Line Engineering with the UML: Deriving Products*, Springer Verlag, 2006, p. 557-586.

Project-Team DYLISS

Dynamics, Logics and Inference for biological Systems and Sequences

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

IN PARTNERSHIP WITH:
CNRS

Université Rennes 1

RESEARCH CENTER
Rennes - Bretagne-Atlantique

THEME
Computational Biology

Table of contents

Project-Team DYLISS

Creation of the Team: 2012 January 01, updated into Project-Team: 2013 July 01

Keywords:

Computer Science and Digital Science:

3.1.1. - Modeling, representation
3.2.3. - Inference
3.2.4. - Semantic Web
3.2.5. - Ontologies
3.3. - Data and knowledge analysis
7.2. - Discrete mathematics, combinatorics
7.3. - Optimization
7.4. - Logic in Computer Science
8.1. - Knowledge
8.2. - Machine learning
8.7. - AI algorithmics

Other Research Topics and Application Domains:

1.1.2. - Molecular biology
1.1.3. - Cellular biology
1.1.9. - Bioinformatics
1.1.11. - Systems biology
1.1.14. - Microbiology
2.2.3. - Cancer
2.2.5. - Immune system diseases

1. Members

Research Scientists

François Coste [Inria, Researcher]
Jacques Nicolas [Inria, Senior Researcher, HDR]
Anne Siegel [Team leader, CNRS, Senior Researcher, HDR]

Faculty Members

Catherine Belleannée [Univ. Rennes I, Associate Professor, 90%-time presence in the lab]
Olivier Dameron [Univ. Rennes I, Associate Professor, HDR]
Laurent Miclet [Univ. Rennes I, Professor, until Sep 2016, HDR]

Technical Staff

Jeanne Got [CNRS]
Meziane Aite [Inria, ADT]
Marie Chevallier [CNRS, ADT until Sept 2016, Regional contract since Oct 2016]
Xavier Garnier [INRA]
Camille Trottier [CNRS, ANR IDEALG]
Pierre Vignet [CNRS, since Nov 2016, ITMO Cancer TGFSysBio]

PhD Students

Aymeric Antoine-Lorquin [Inria]

Lucas Bourneuf [Univ. Rennes I, since Oct 2016]
Jean Coquet [Univ. Rennes I]
Victorien Delannée [Univ. Rennes I]
Clémence Frioux [Inria]
Julie Laniau [Inria]
Yann Rivault [Univ. Rennes I]

Post-Doctoral Fellow

Aurélie Evrard [INRA, until Aug 2016]

Visiting Scientists

Oumarou Abdou Arbi [Univ. of Maradi - Niger, from Jan 2016 until Feb 2016]
Maria Paz Cortes Burgos [Univ. of Chile, from may 2016 until Nov 2016]
Mauricio Alejandro Latorre [Univ. of Chile, Sep 2016]
Alejandro Maass [Univ. of Chile, Sep 2016]
Santiago Videla [Institut Leloir, Argentina, Jul 2016]

Administrative Assistant

Marie Le Roic [Univ. Rennes I]

Others

Jérémie Bourdon [Univ. Nantes, Dyliss associate member, 20%-time presence in the lab, HDR]
Damien Eveillard [Univ. Nantes, Dyliss associate member, 20%-time presence in the lab]
François Moreews [INRA, Dyliss partner, 80%-time presence in the lab]
Nathalie Théret [INSERM, Senior Researcher, Dyliss partner, 80%-time presence in the lab, HDR]
Denis Tagu [INRA, Senior Researcher, Dyliss associate member, 20%-time presence in the lab, HDR]

2. Overall Objectives

2.1. Overall objectives

The research domain of the bioinformatics Dyliss team is sequence analysis and systems biology. Our main goal in biology is to characterize groups of genetic actors that control the phenotypic answer of species when challenged by their environment. The team explores methods in the field of formal systems, more precisely in knowledge representation, constraints programming, multi-scale analysis of dynamical systems, and machine learning. Our goal is to identify key regulators of the environmental response by structuring and reasoning on information which combines physiological responses measured with omics technologies (RNA-seq, metabolomics, proteomics), genetic information from their long-distant cousins and knowledge about regulation and metabolic pathways stored in public repositories.

The main challenges we face are data incompleteness and heterogeneity. We favor the construction and study of a "space of feasible models or hypotheses" including known constraints and facts on a living system *rather than searching for a single optimized model*. We develop methods allowing a precise investigation of this space of hypotheses. Therefore, we are in position of developing experimental strategies to progressively shrink the space of hypotheses and gain in the understanding of the system. Importantly, one should notice that our models spans a quite large spectrum of discrete structures: oriented graphs, boolean networks, automata, or expressive grammars.

More concretely, the steps of the analysis are to (*i*) formalize and integrate in a set of logical or grammatical constraints both generic knowledge information (litterature-based regulatory pathways, diversity of molecular functions, DNA patterns associated with molecular mechanisms) and species-specific information (physiological response to perturbation, sequencing...); (*ii*) investigate the space of admissible models and exhibit its main features by solving combinatorial optimization problems; (*iii*) identify corresponding genomic products within sequences. At each of these steps, we rely on symbolic methods for model space exploration: ontologies and formal concepts analysis.

We target applications for which large-scale heterogeneous data about a specific but complex physiological phenotype are available. Existing long-term partnerships with biological labs give strong support to this choice. In marine biology, we collaborate closely with the Station biologique de Roscoff (*Idealg*, Investissement avenir "Bioressources et Biotechnologies"). In environmental microbiology we collaborate both with the CRG in Chile in the framework of the Ciric Chilean Inria center (*Ciric-Omics*). In agriculture, our main partners are within the INRA institute in Rennes, with a focus on the understanding of pea-aphids microbiology and of breeding animals metabolism (porc, chicken, cow). More recently, we have introduced health as a new application field of the team, especially through the study of large-scale boolean networks and their confrontation with knowledge repositories.

3. Research Program

3.1. Modeling knowledge integration with combinatorial constraints

Biological networks are built with data-driven approaches aiming at translating genomic information into a functional map. Most methods are based on a probabilistic framework which defines a probability distribution over the set of models. The reconstructed network is then defined as the most likely model given the data.

Our team has investigated an alternative perspective where each data induces a set of constraints - related to the steady state response of the system dynamics - on the set of possible values in a network of fixed topology. The methods that we have developed complete the network with product states at the level of nodes and influence types at the level of edges, able to globally explain experimental data. In other words, the selection of relevant information in the model is no more performed by selecting *the* network with the highest score, but rather by exploring the complete space of models satisfying constraints on the possible dynamics supported by prior knowledge and observations. In the (common) case when there is no model satisfying all the constraints, we relax the problem by introducing new combinatorial optimization problems that introduce the possibility of correcting the data or the knowledge. Common properties to all solutions are considered as a robust information about the system, as they are independent from the choice of a single solution to the optimization problem [6].

Solving these computational issues requires addressing NP-hard qualitative (non-temporal) issues. We have developed a long-term collaboration with Potsdam University in order to use a logical paradigm named **Answer Set Programming** (ASP) [50], [69] to solve these constraint satisfiability and combinatorial optimization issues. Applied on transcriptomic or cancer networks, our methods identified which regions of a large-scale network shall be corrected [51], and proposed robust corrections [5]. This result suggested that this approach was compatible with efficiency, scale and expressivity needed by biological systems.

During the last years, our goal was to provide **formal models of queries on biological networks** with the focus of integrating dynamical information as explicit logical constraints in the modeling process. Using these technologies requires to revisit and reformulate constraint-satisfiability problems at hand in order both to decrease the search space size in the grounding part of the process and to improve the exploration of this search space in the solving part of the process. Concretely, getting logical encoding for the optimization problems forces to clarify the roles and dependencies between parameters involved in the problem. This paves the way to a refinement approach based on a fine investigation of the space of hypotheses in order to make it smaller and gain in the understanding of the system. Our studies confirmed that logical paradigms are a powerful approach to build and query reconstructed biological systems, in complement to discriminative ("black-box") approaches based on statistical machine-learning. Based on these technologies, we have developed a panel of methods allowing the integration of muli-scale data knowledge, linking genomics, metabolomics, expression data and protein measurement of several phenotypes (see Fig. 1).

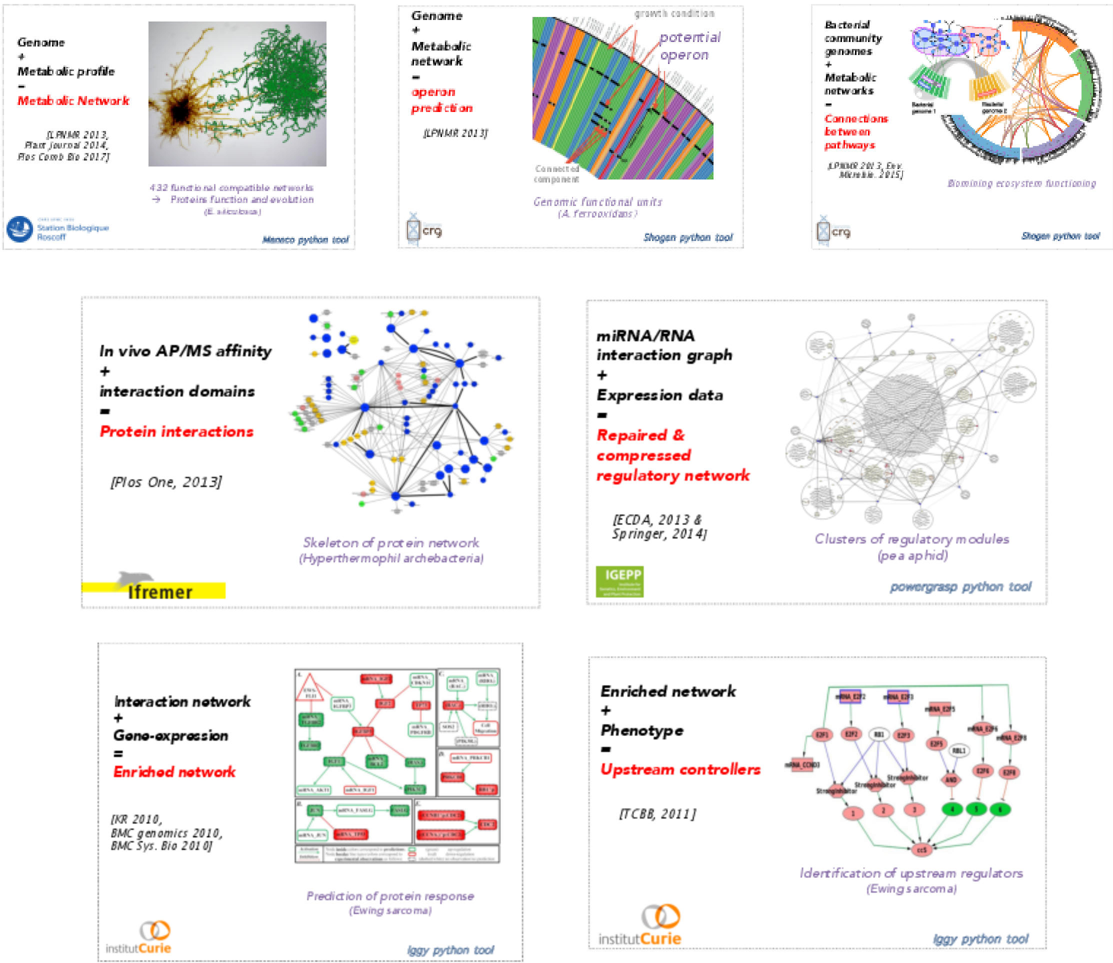


*Figure 1. **Multi-scale data and knowledge integration procedures** Dynamical analyses are undergone to elucidate relationships between biological scales. Such dependencies are combined with data and turn into a first order logic paradigm (Answer Set Programming). Key interactions or genes of interest are then identified to be the solution of a combinatorial optimization problem. Methods are encapsulated in python packages and provided in the meta-package bioasp.*

Notice that our main issue is in the field of knowledge representation. More precisely, we do not wish to develop new solvers or grounders, a self-contained computational issue which is addressed by specialized teams such as our collaborator team in Potsdam. Our goal is rather to investigate how the constant progresses in the field of constraint logical programming, shown by the performance of ASP-solvers, are sufficient to address the complexity of constraint-satisfiability and combinatorial optimization issues explored in systems biology. In this direction, we work in close interaction with Potsdam university to feed their research activities which challenging issues from bioinformatics and, as a feed-back, take benefit of the prototypes they develop.

By exploring the complete space of models, our approach typically produces numerous candidate models compatible with the observations. We began investigating to what extent domain knowledge can further refine the analysis of the set of models by identifying classes of similar models, or by selecting a subset of models that satify an additional constraint (for instance, best fit with a set of experiments, or with a minimal size). We anticipate that this will be particularly relevant when studying non-model species for which little is known but valuable information from other species can be transposed or adapted. These efforts consist in developing reasoning methods based on ontologies as formal representation of symbolic knowledge. We use Semantic Web tools such as SPARQL for querying and integrating large sources of external knowledge, and measures of semantic similarity and particularity for analyzing data.

3.2. Modeling the dynamical response of biological systems with logical and (non)-linear constraints

As explained below, Answer Set programming technologies enable the identification of key controllers based on the integration of static data. As a natural follow-up, we also develop optimization techniques to learn models of the dynamics of a biological system. As before, our strategy is not to select a single model fitting with experimental data but rather to decipher the complete set of families of models which a compatible with the observed response. Our main research line in this field is to decipher the appropriate level of expressivity (in terms of constraints) allowing both to properly report the nature of data and knowledge and to allow for an exhaustive study of the space of feasible models. To implement this strategy, we rely on several constraint programming frameworks, which depend on the model scale and the nature of time-points kinetic measurements. The three following examples are shown in Fig. 2.

- In [7], logical programming (Answer Set programming) is used to decipher the combinatorics ot synchrone boolean networks explaining static or dynamics response of signaling networks to perturbations (such as measured by phosphoproteomics technologies).
- In [49], SAT-based approaches are used to decipher the combinatorics of large-scale asynchronous boolean networks. In order to gain in expressivity, we model these networks as guarded-transition network, an extension of Petri nets.
- In [2] and [47], linear Programming frameworks are used to decipher the variability of the response of reaction-based networks. Still to gain in expressivity, we model systems with Markovian qualitative description of its dynamics together with quantitative laws which describe the effect of the dynamic transitions over higher scale quantitative measurements. Families of models are investigated with ad-hoc local search algorithms.
- Finally, classical learning methods are used to build ad-hoc parameterized numerical models that provide the most parsimonious explanations to experimental measurements.

3.3. Modeling sequences with formal grammars

Once groups of genome products implied in the answer of the species have been identified with integrative or dynamics methods, it remains to characterize the biological actors within genomes. To that goal, we both learn, model and parse formal patterns within DNA, RNA or protein sequences. More precisely, our research on modeling biomolecular sequences with expressive formal grammars focuses on learning such grammars from examples, helping biologists to design their own grammar and providing practical parsing tools.

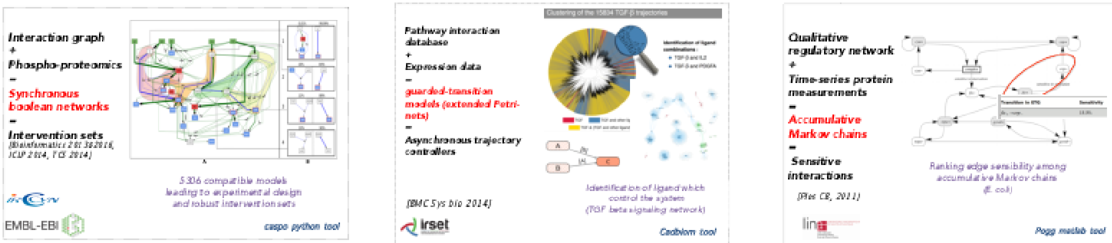


*Figure 2. **Learning and investigating complete families of dynamical models compatible with available data.** Depending on the scale of the system and the nature of data, we use synchronous boolean networks, enriched Petri Nets or accumulative Markov chains to report and explain the measured response of a biological systems.*

On the development of **machine learning** algorithms for the induction of grammatical models [40], we have a strong expertise on learning finite state automata. We have proposed an algorithm that learns successfully automata modeling families of (non homologous) functional families of proteins [4], leading to a tool named Protomata-learner (see Fig. 3). The algorthim is based on a similar fragment merging heuristic approach which reports partial and local alignments contained in a family of sequences.. As an example, this tool allowed us to properly model the TNF protein family, a difficult task for classical probabilistic-based approaches. It was also applied successfully to model important enzymatic families of proteins in cyanobacteria [3]. Our future goal is to further demonstrate the relevance of formal language modeling by addressing the question of a fully automatic prediction from the sequence of all the enzymatic families, aiming at improving even more the sensitivity and specificity of the models. As enzyme-substrate interactions are very specific central relations for integrated genome/metabolome studies and are characterized by faint signatures, we shall rely on models for active sites involved in cellular regulation or catalysis mechanisms. This requires to build models gathering both structural and sequence information in order to describe (potentially nested or crossing) long-term dependencies such as contacts of amino-acids that are far in the sequence but close in the 3D protein folding. Our current researches is focused on the inference of Context-Free Grammars including the topological information coming from the structural characterization of active sites.

Using context-free grammars instead of regular patterns increases the complexity of **parsing** issues. Indeed, efficient parsing tools have been developed to identify patterns within genomes but most of them are restricted to simple regular patterns. Definite Clause Grammars (DCG), a particular form of logical context-free grammars have been used in various works to model DNA sequence features [76]. An extended formalism, String Variable Grammars (SVGs), introduces variables that can be associated to a string during a pattern search (see Fig. 3) [90], [89]. This increases the expressivity of the formalism towards mildly context sensitive grammars. Thus, those grammars model not only DNA/RNA sequence features but also structural features such as repeats, palindromes, stem/loop or pseudo-knots. Few years ago, we have designed a first tool, STAN (suffix-tree analyser), in order to make it possible to search for a subset of SVG patterns in full chromosome sequences [8]. This tool was used for the recognition of transposable elements in *Arabidopsis thaliana* [92]. We have enlarged this experience through a new modeling language, called Logol [1]. Generally, a suitable language for the search of particular components in languages has to meet several needs : expressing existing structures in a compact way, using existing databases of motifs, helping the description of interacting components. In other words, the difficulty is to find a good tradeoff between expressivity and complexity to allow the specification of realistic models at genome scale. The Logol language and associated framework have been built in this direction. See Figure 3 for illustration. The Logol specificity beside other SVG-like languages mainly lies in a systematic introduction of constraints on string variables.

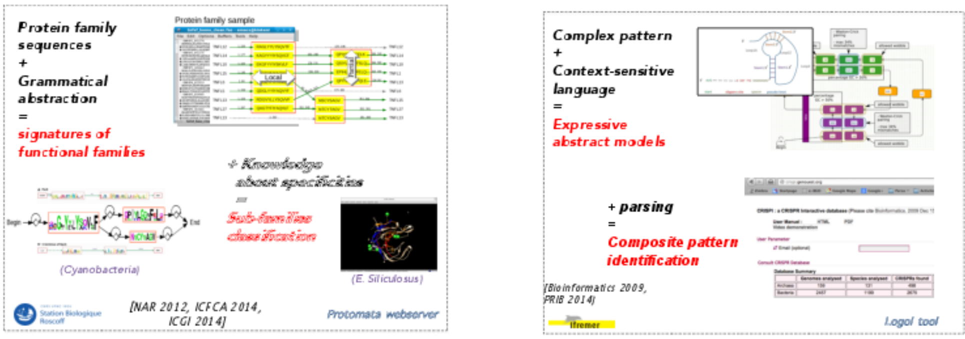


*Figure 3. **Learning and parsing sequences of genome or protein families with expressive grammars**. (a) The protomata workflow starts from a set of protein sequences. A partial local alignment is computed and an automaton is inferred, which can be considered as a signature of the family of proteins. This allows searching for new members of the family [3]. Adding further information about the specific properties of proteins within the family allows to exhibit a refined classification. (b) The Logol framework allows modeling complex structure in sequences, such as a pseudo-knot (RNA structure). This is based on the expressivity of String Variable Grammars. Combined with parsers, this leads to composite pattern identification such as CRISPR. [84].*

3.4. Symbolic methods for model space exploration: Semantic web for life sciences and Formal Concepts Analysis

All the methods presented in the previous sections usually result in pools of candidates which equivalently explain the data and knowledge. These candidates can be dynamical systems, compounds, biological sequences, proteins... In any case, the output of our formal methods generally requires *a posteriori* investigation and filtering by domain experts. In order to assist them, we rely on two classes of symbolic technics: Semantic Web technologies and Formal Concept Analysis (FCA). They both aim at the formalization and management of knowledge, that is, the explicitation of relations occuring in structured data. These technics complement each other: the production of relevant concepts in FCA highly depends on the availability of semantic annotations using a controlled set of terms and conversely, building and exploiting ontologies is a complex process that can be made much easier with FCA.

Integrating heterogenous data with semantic web technologies The emergence of ontologies in biomedical informatics and bioinformatics happened in parallel with the development of the **Semantic Web** in the computer science community [88]. Let us recall that the Semantic Web is an extension of the current Web that provides an infrastructure integrating data and ontologies in order to support unified reasoning. Since the beginning, life sciences have been a major application domain for the Semantic Web [52]. This was motivated by the joint evolution of data acquisition capabilities in the biomedical field, and of the methods and infrastructures supporting data analysis (grids, the Internet...), resulting in an explosion of data production in complementary domains [60], [53]. Consequently, Semantic Web technologies have become an integral part of translational medicine and translational bioinformatics [63]. The Linked Open Data project promotes the integration of data sources in machine-processable formats compatible with the Semantic Web [59], with a strong involvement of life sciences in this initiative.

However, a specificity of life sciences "data deluge" is that the proportion of generated data is much higher than in the more general "big data phenomenon", and that these data are highly connected [91]. **The bottleneck that once was data scarcity now lies in the lack of adequate methods supporting data integration, processing and analysis**. [78]. Each of these steps typically hinges on domain knowledge, which is why they resist automation. This knowledge can be seen as the set of rules representing in what conditions data can be used or can be combined for inferring new data or new links between data.

In this setting, we are working on the integration of Semantic Web resources with our data analysis methods in order to take existing biological knowledge into account. We have introduced several methods to interpret semantic similarities and particularities [58], [57]. We now focus our attention on the semi-automated construction of RDF abstractions of heterogeneous datasets which can be handled by non-expert users. This allows both to automatically prepare input datasets for the other methods developed in the team and to analyse the output of the methods in a wide knowledge context.

*Figure 4. **Data-sciences methods based-on semantic-web technologies and formal concept analysis** allows for the knowledge-based post-processing of the results of bioinformatics methods.*

Using Formal concept analysis to explore the results of bioinformatics analyses Formal concept analysis aims at the development of conceptual structures which can be logically activated for the formation of judgments and conclusions [96]. It is used in various domains managing structured data such as knowledge processing, information retrieval or classification [79]. In its most simple form, one considers a binary relation between a set of objects and a set of attributes. In this setting, formal concept analysis formalizes the semantic notions of extension and intension. Concepts are related within a lattice structure (Galois connection) by subconcept-superconcept relations, and this allows drawing causality relations between attribute subsets. In bioinformatics, it has been used to derive phylogenetic relations among groups of organisms [77], a classification task that requires to take into account many-valued Galois connections. We have proposed in a similar way a classification scheme for the problem of protein assignment in a set of protein families [65].

One of the most important issue with concept analysis is due to the fact that current methods remain very sensitive to the presence of uncertainty or incompleteness in data. On the other hand, this apparent defect can be reversed to serve as a marker of incompleteness or inconsistency [66]. Following this inspiration, we have proposed a methodology to tackle the problem of uncertainty on biological networks where edges are mostly predicted links with a high level of false positives [97]. The general idea consists to look for a tradeoff between the simplicity of the conceptual representation and the need to manage exceptions. As a very prospective challenge, we are exploring the idea of using ontologies to help this or to help ontology refinement using concept analysis [80], [56], [83].

More generally, common difficult tasks in this context are visualization, search for local structures (graph mining) and network comparison. Network compression is a good solution for an efficient treatment of all these tasks. This has been used with success in power graphs, which are abstract graphs where nodes are clusters of nodes in the initial graph and edges represent bicliques between two sets of nodes [85]. In fact, concepts are maximal bicliques and we are currently developing the power graph idea in the framework of concept analysis.

4. Application Domains

4.1. Application domain in bioinformatics

As mentioned before, our main goal in biology is to characterize groups of genetic actors that control the response of living species capable of facing extreme environments. To focus our developments, applications and collaborations, we have identified three biological questions which deserve integrative studies. Each axis may be considered independently from the others although their combination, a mid-term challenge, will have the best impact in practice towards the long-term perspective of identifying proteins controlling the production of a metabolite of industrial interest. It is illustrated in our presentation for a major algae product: polyunsaturated fatty acids (PUFAs) and their derivatives.

Integrative biology with combinatorial optimization. The first axis of the project (data integration) aims at identifying *who* is involved in the specific response of a biological system to an environmental stress. Targeted actors will mainly consist in groups of genetic products or biological pathways. For instance, which pathways are involved in the specific production of PUFAs in brown algae? The main work is to represent in a system of logical constraints the full knowledge at hand concerning the genetic or metabolic actors, the available observations and the effects of the system dynamics. To this aim, we focus on the use of Answer Set Programming as we are experienced in modeling with this paradigm and we have a strong partnership with a computer science team leader in the development of dedicated grounders and solvers (Potsdam university). See Sec. 3.1.

Systems biology with discrete dynamical modeling. Once a model is built and its main actors are identified, the next step is to clarify *how* they combine to control the system. This is the second axis of the project. Roughly, the fine tuning of the system response may be of two types. Either it results from the discrete combinatorics of the actors, as the result of a genetic adaptation to extreme environmental conditions or the difference between species is rather at the enzyme-efficiency level. For instance, if PUFAs are found to be produced using a set of pathways specific to brown algae, our work on dynamical modeling will consist to apply constraint-based combinatorial approaches to select consistent combinations of pathways controlling the metabolite production. Otherwise, if enzymes controlling the production of PUFAs are found to be expressed in other algaes, it suggests that the response of the system is rather governed by a fine quantitative tuning of pathways. In this case, we use symbolic dynamics and average-case analysis of algorithms to weight the respective importance of interactions in observed phenotypes (see Sec. 3.2 and Fig. 2). This specific approach is motivated by the quite restricted spectrum of available physiological observations over the asymptotic dynamics of the biological system.

Biological sequence annotation with grammatical inference and modelling In order to check the accuracy of in-silico predictions, a third research axis of the team is to extract genetic actors responsible of biological pathways of interest in the targeted organism and locate them in the genome. In our guiding example, active proteins implied in PUFAs controlling pathways have to be precisely identified. Actors structures are represented by syntactic models (see Fig. 3). We use knowledge-based induction on far instances for the recognition of new members of a given sequence family within non-model genomes (see Fig. 3). A main objective is to model enzyme specificity with highly expressive syntactic structures - context-free model - in order to take into account constraints imposed by local domains or long-distance interactions within a protein sequence. See Sec. 3.3 for details.

Data classification with data sciences All the methods presented in the previous section usually result in pools of candidates which equivalently explain the data and knowlegde. These candidates can by dynamical systems, compounds, biological sequences, proteins... In any case, the output of our formal methods generally deserves a a-posteriori investigation and filtering. To that goal, we rely on two classes of symbolic technics: semantic web technologies and Formal Concept Analysis See Sec. 3.4 for details.

4.2. Application fields in biology

Our methods are applied in several fields of molecular biology.

Our main application field is **marine biology**, as it is a transversal field with respect to issues in integrative biology, dynamical systems and sequence analysis. Our main collaborators work at the Station Biologique de Roscoff. We are strongly involved in the study of brown algae: the *meneco, memap and memerge* tools were designed to realize a complete reconstruction of metabolic networks for non-benchmark species [82], [68]. On the same application model, the pattern discovery tool *protomata learner* combined with supervised bi-clustering based on formal concept analysis allows for the classification of sub-families of specific proteins [65]. The same tool also allowed us to gain a better understanding of cyanobacteria proteins [3]. At the larger level of 4D structures, classification technics have also allowed us to introduce new methods for the characterization of viruses in marine metagenomic sample [19]. Finally, in dynamical systems, we use asymptotic analysis (tool *pogg*) to decipher the initiation of sea urchin translation [55] [24]. We are currently involved in two new applications in this domain: the team participates to a Inria Project Lab program with the Biocore and Ange Inria teams, focused on the understanding on green micro-algae; and we are involved in the deciphering of phytoplancton variability at the system biology level in collaboration with the Station Biologique de Roscoff (ANR Samosa).

In **micro-biology**, our main issue is the understanding of bacteria living in extreme environments, mainly in collaboration with the group of bioinformatics at Universidad de Chile (funded by CMM, CRG and Inria-Chile). In order to elucidate the main characteristics of these bacteria, we develop efficient methods to identify the main groups of regulators for their specific response in their living environment. To that purpose, we use constraints-based modeling and combinatorial optimization. The integrative biology tools *meneco bioquali, ingranalysis, shogen, lombarde* were designed in this context [6]. in 2016, two applications focused on the study of extremophile consortium of bacteria have been performed with these tools [15], [13]. In parallel, in collaboration with Ifremer (Brest), we have conducted similar work to decipher protein-protein interactions within archebacteria [81]. Our sequence analysis tool (*logol*) allowed us to build and maintain a very expressive CRISPR database [10] [54].

Similarly, in **agriculture**, our goal is to propose methods to identify regulators of very complex phenotypes related to environmental issues. In collaboration with researchers from Inra/Pegase and Inra/Igeep laboratories, we develop methods to distinguish the response of breeding animals to different diaries or treatments [47] and characterize upstream transcriptional regulators [61], with applications in porks [70], [71] [20]. The pattern matching tool *logol* also allows for a fine identification of transcription factor motifs applied to chicken [67] [54]. Semantic-based analysis was useful for interpreting differences of gene expression in pork meat [72]. Finally, Constraints-based programming also allows us to decipher regulators of reproduction for pea aphids [75], [98] and paved the way to the recent research track initiated in the team about integration of heterogeneous data with RDF-technologies (see askomics software) [37], [45].

Similarly, in **agriculture**, our goal is to propose methods to identify regulators of very complex phenotypes related to environmental issues. In collaboration with researchers from Inra/Pegase laboratory, we develop methods to distinguish the response of breeding animals to different diaries or treatments [47] and characterize upstream transcriptional regulators [61], applied to porks [70], [71] [20]. The pattern matching tool *logol* also allows for a fine identification of transcription factor motifs applied to chicken [67] [54]. Semantic-based analysis was useful for interpreting differences of gene expression in pork meat [72].

In addition, constraints-based programming also allows us to decipher regulators of reproduction for the pea aphid, an insect that is a pest on plants [75], [98]. This was performed in collaboration with Inra/Igepp . This paved the way to the recent research track initiated in the team about integration of heterogeneous data with RDF-technologies (see askomics software) [37], [45] and about graph-compression (see powergrasp software).

In **bio-medical applications**, we focus our attention on the confrontation of large-scale measurements with large-scale knowledge repositories about regulation pathways such as Transpath, PID or pathway commons. In collaboration with Institut Curie, we have studied the Ewing Sarcoma regulation network to test the capability of our tool *bioquali* to accurately correct and predict a large-scale network behavior [51]. Our ongoing studies in this field focus on the exhaustive learning of discrete dynamical networks matching with experimental data, as a case study for modeling experimental design with constraints-based approaches. To that purpose, we collaborate with J. Saez Rodriguez group at EBI [94] and N. Theret group at Inserm/Irset (Rennes) [49]. The dynamical system tools *caspo and cadbiom* were designed within these collaborations. Ongoing studies focus on the understanding of the metabolism of xenobiotics (mecagenotox program) and the filtering of sets of regulatory compounds within large-scale signaling network (TGFSysBio project).

5. Highlights of the Year

5.1. Highlights of the Year

The first main novelty in 2016 is the release of our first methods and tools based on semantic web technologies. These methods enable the pre-processing of heterogeneous data prior to their integration in the toolboxes developped by the team. Methods for the transparent integration and querying of heterogeneous data (AskOmics) as well as the user-friendly tracable reconstruction of metabolic networks (PADmet package) have been developped in collaboration with our main partners (INRA Rennes, University of Chile, Station biologique de Roscoff) to facilitate the comparison of phenotypes accross several species or several strains.

6. New Software and Platforms

6.1. AskOmics

KEYWORDS: RDF - SPARQL - Querying - Graph
FUNCTIONAL DESCRIPTION

AskOmics allows to load heterogeneous bioinformatics data (formatted as tabular files or directly in RDF) into a Triple Store system using a user-friendly web interface. AskOmics also provides an intuitive graph-based user interface supporting the creation of complex queries that currently require hours of manual searches across tens of spreadsheet files. The elements of interest selected in the graph are then automatically converted into a SPARQL query that is executed on the users' data.

- Authors: Charles Bettembourg, Yvanne Chaussin, Anthony Bretaudeau, Olivier Filangi, Fabrice Legeai and Olivier Dameron
- Partners: CNRS - INRA - Université de Rennes 1
- Contact: Fabrice Legeai
- https://github.com/askomics/askomics

6.2. PADMet

PortAble Database for Metabolism
KEYWORDS: Bioinformatics - Toolbox - Metabolic networks - Standardization
FUNCTIONAL DESCRIPTION

The PADMet package allows conciliating genomics and metabolic network information used to produce a genome-scale constraint-based metabolic model within a database that traces all the reconstruction process steps. It allows representing the metabolic model in the form of a Wiki containing all the used/traced information. Other standard outputs are made available with the package. The main concept underlying PADMet-Package is to provide solutions that ensure the consistency, the internal standardization and the reconciliation of the information used within any workflow that combines several tools involving metabolic networks reconstruction or analysis. The PADMet package is at the core of the AuReMe workflow, dedicated to the primary reconstruction of genome-scale metabolic networks from raw data. It allows the study of organisms for which few experimental data are available. Its main feature is to undergo the reconstruction of the metabolic network by combining several heterogeneous knowledge and data sources, including the information reported by several scaffold metabolic networks for cousin species.

- Partners: CNRS - Inria - Université de Rennes 1 - University of Chile.
- Contact: Meziane Aite
- https://gitlab.inria.fr/DYLISS/padmet-toolbox

6.3. PowerGrASP

Power Graph compression in ASP

KEYWORDS: Bioinformatics - Constraint-based Programming - Data visualization - Optimization - Decomposition - Graph - Graph visualization - Pattern extraction - Answer Set Programming - Formal concept analysis

FUNCTIONAL DESCRIPTION

Implementation of graph compression methods oriented toward visualization, and based on power graph analysis. The method relies of formal concept analysis and is implemented in the declarative langage Answer Set Programming. It is applied to regulatory networks currently produced in the domain of bioinformatics.

- Participants: Lucas Bourneuf, Jacques Nicolas
- Partners: Inria - Université de Rennes 1 - INRA.
- Contact: Lucas Bourneuf
- URL: http://github.com/aluriak/powergrasp

6.4. Platforms and toolboxes

A goal of the team is to facilitate interplays between tools for biological data analysis and integration. Our tools aim at guiding the user to progressively reduce the space of models (families of sequences of genes or proteins, families of keys actors involved in a system response, dynamical models) which are compatible with both knowledge and experimental observations.

Most of our tools are developed in collaboration with the GenOuest resource and data center hosted in the IRISA laboratory, including their computer facilities [more info]. It worths considering them into larger dedicated environments to benefit from the expertise of other research groups.

- The **BioShadock** repository allows one to share the different docker containers that we are developing [website].
- The **Inria chile Mobyle portal** gathers some of the tools that were developed in collaboration with Dyliss, such as meneco, shogen and lombarde [website].
- The **bioASP portal** gather most of ASP-based python packages that we are developping in collaboration with Potsdam university [website]
- The **GenOuest galaxy portal** now provides access to most tools for integrative biology and sequence annotation (access on demand).

6.4.1. AuReMe - Tracable reconstruction of metabolic networks

The toolbox **AuReMe** allows for the **Au**tomatic **Re**construction of **Me**tabolic networks based on the combination of multiple heteregeneous data and knowledge sources. Since 2016, the workflow has been made available as a Docker image to facilitate its distribution among the scientific community [web page].

- The **Model-management PADmet module** allows conciliating genomics and metabolic network information used to produce the metabolic model within a local database that traces all the reconstruction process steps and to connect software in the pipeline. This toolbox was completly redesigned in 2016. [package]
- The **meneco python package** allows filling the gaps of a metabolic network by using a qualitative approach to elaborate the biosynthetic capacities; the problem is viewed as a combinatorial optimization problem encoded in a Answer Set Programming Problem [87] [64]. [python package].
- The **shogen python package** allows aligning genome and metabolic network to identify genome units which contain a large density of genes coding for enzymes that regulate successive reactions of metabolic pathways; the problem is also encoded with an ASP program. [62]. [python package].
- The **Manual curation assistance PADmet module** allows for curating the reported metabolic networks and modify metadata [package].
- The **Wiki-export PADmet module** enables the export of the metabolic network and its functional genomic unit as a local wiki platform allowing the user-friendly investigation of the network together with the main steps used to reconstruct it. It was developed in 2016. [package].

6.4.2. Filtering interaction networks with graph-based optimization criteria

The goal is to offer a toolbox for the reconstruction of networks from genome, literature and large-scale observation data (expression data, metabolomics...) in order to elucidate the main regulators of an observed phenotype. Most of the optimization issues are addressed with Answer Set Programming.

- The **lombarde package** enables the filtering of transcription-factor/binding-site regulatory networks with mutual information reported by the response to environmental perturbations. The high level of false-positive interactions is filters according to graph-based criteria. Knowledge about regulatory modules such as operons or the output of the shogen package can be taken into account [48][13] [web server].
- The **KeyRegulatorFinder package** allows searching key regulators of lists of molecules (like metabolites, enzymes or genes) by taking advantage of knowledge databases in cell metabolism and signaling. The complete information is transcribed into a large-scale interaction graph which is filtered to report the most significant upstream regulators of the considered list of molecules [61] [package].
- The **powerGrasp python package** provides an implementation of graph compression methods oriented toward visualization, and based on power graph analysis. [package].
- The **iggy package** enables the repairing of an interaction graph with respect to expression data. It proposes a range of different operations for altering experimental data and/or a biological network in order to re-establish their mutual consistency, an indispensable prerequisite for automated prediction. For accomplishing repair and prediction, we take advantage of the distinguished modeling and reasoning capacities of Answer Set Programming. [5] [93] [Python package][web server].

6.4.3. Caspo - Studying synchronous boolean networks

The **caspo** pipeline is dedicated to automated reasoning on logical signaling networks. The main underlying issue is that inherent experimental noise is considered, so that many different logical networks can be compatible with a set of experimental observations.

Software provides an easy to use software for the study of synchronous logical (boolean) networks. In 2016, the tool was redesigned to enhance its functionalities and integrated in a docker container to facilitate its use on any platform [86] [28] [python package and docker container].

- The **caspo-learn module** performs an automated inference of logical networks from the observed response to different perturbations (phosphoproteomics datasets). It allows for identifying admissible large-scale logic models saving a lot of efforts and without any a priori bias. It is also included in the cellNopt package [0] [7] [94].
- The **caspo-classify, predict and visualize modules** allows for classifying a family of boolean networks with respect to their input–out- put predictions [7].
- The **caspo-design module** designs experimental perturbations which would allow for an optimal discrimination of rival models in a family of boolean networks [95].
- The **caspo-control module** identifies key-players of a family of networks: it computes robust intervention strategies (i.e. inclusion minimal sets of knock-ins and knock-outs) that force a set of target species or compounds into a desired steady state [73].
- **caspo-timeseries module** have been designed by our colleagues from LRI as an extension of the caspo pipeline to take into account time-series observation datasets in the learning procedure [23] [python package and docker container].

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

6.4.4. cadbiom - Building and analyzing the asynchronous dynamics of enriched logical networks

Based on Guarded transition semantic, the **cadbiom** software provides a formal framework to help the modeling of biological systems such as cell signaling network. It allows investigating synchronization events in biological networks. In 2016, the tool was integrated in a docker container in order to facilitate its use on any platform [49] [docker container][web server].

- **The cadbiom graphical interface** is useful to build and study moderate size models. It provides means for model exploration, simulation and checking. For large-scale models, the graphical interface allows to focus on specific nodes of interest.
- **The cadbiom API** allows to load a model (including large-scale ones), perform static analysis (exploration, frontier computation, statistics, and dependence graph computation) and check temporal properties on a finite horizon in the future or in the past.
- **Exploring large-scale knowledge repositories** A main feature of cadbiom is that automatic translation of the large-scale PID repository (about 10,000 curated interactions) have been automatically translated into the cadbiom formalism. Therefore, the API allows for computing the upstream regulators of any set of genes based on this large-scale repository.

6.4.5. Protomata - Expressive pattern discovery on protein sequences

Protomata is a machine learning suite for the inference of *automata* characterizing (functional) families of proteins from available sequences. Based on a new kind of alignment said partial and local, it learns precise characterizations of the families – beyond the scope of classical sequence patterns such as PSSM, Profile HMM, or Prosite Patterns – allowing to predict new family members with a high specificity.

Protomata gives access to the three main modules as stand-alone programs, which are also integrated in a single workflow *protomata-learner*:

- **Paloma** builds partial local multiple alignments;
- **Protobuild** infers automata from these alignements;
- **Protomatch and Protoalign** scan, parse and align new sequences based on the automata infered previously. This module was improved in 2016 by embedding new options to score the sequences with respect to all accepting paths (Forward score) in addition to the scoring module based on the best path (Viterbi score). More generally, we have worked on the efficiency of the automata's weighting scheme based on the state-of-the-art schemes used for profile HMMs.

The suite is completed by many tools to handle or visualize data and can be used online via a [web interface].

6.4.6. Logol - Complex pattern modelling and matching

The **Logol** toolbox is a swiss-army-knife for pattern matching on DNA/RNA/Protein sequences, using a high-level grammatical formalism to permit a large expressivity for patterns [54]. A Logol pattern consists in a complex combination of motifs (such as degenerated strings) and structures (such as imperfect stem-loop ou repeats). Compared to other specialized pattern matching tools, some of the Logol key features are the possibilities to divide a pattern description into several sub-patterns, to enable the use of ambiguous models or to permit the inclusion of negative conditions in a pattern definition. Possible fields of application are the detection of mutated binding sites [32] or stem-loop identification (e.g. in CRISPR [0] [10]) [web page].

- The **Graphical designer** allows a user to iteratively build a complex pattern based on basic graphical patterns. The associated grammar file is an export of the graphical designer. In 2015, the efficiency of the tool was improved by slight evolutions of the underlying grammar.
- The **LogolMatch** parser takes as input a biological (nucleic or amino acid) sequence and a grammar file (i.e. a pattern). It combines a grammar analyzer, a sequence analyzer and a prolog Library. It returns a file containing all the occurrences of the pattern in the sequence with their parsing details.
- Full genome analysis, and connection to biological databases have been made available recently.

[0] http://crispi.genouest.org/

7. New Results

7.1. Data integration and pre-processing with semantic-based technologies

Participants: Meziane Aite, Marie Chevallier, Olivier Dameron, Aurélie Evrard, Clémence Frioux, Xavier Garnier, Jeanne Got, François Moreews, Yann Rivault, Anne Siegel, Pierre Vignet, Denis Tagu, Camille Trottier.

Integration and query of biological datasets with Semantic Web technologies. The purpose of this work is to obtain quick answers to biological questions demanding currently hours of manual search in several spreadsheet results files. We introduce an integration and interogation framework using an RDF model and the SPARQL query language. It allows biologists to transparently integrate and query their data without any a priori proficiency about RDF and SPARQL. [*O. Dameron, A. Evrard, X. Garnier*] [37], [45]

Handling the heterogeneity of genomic and metabolic networks data within flexible workflows with the PADMet toolbox A main challenge of the era of fast and massive genome sequencing is to transform sequences into biological knowledge. The high diversity of input files and tools required to run any metabolic networks reconstruction protocol represents an important drawback: it appears very difficult to ensure that input files agree among them. Such a heterogeneity produces loss of information during the use of the protocols and generates uncertainty in the final metabolic model. Here we introduce the PADMet-toolbox which allows conciliating genomic and metabolic network information. The toolbox centralizes all this information in a new graph-based format: PADMet (PortAble Database for Metabolism) and provides methods to import, update and export information. For the sake of illustration, the toolbox was used to create a workflow, named AuReMe, aiming to produce high-quality genome-scale metabolic networks and eventually input files to feed most platforms involved in metabolic network analyses. We applied this approach to two exotic organisms and our results evidenced the need of combining approaches and reconciling information to obtain a functional metabolic network to produce biomass. [*M. Chevallier, M. Aite, C. Frioux, J. Got, A. Siegel, C. Trottier, P. Vignet*] [34]

PEPS: a platform for supporting studies in pharmaco-epidemiology using medico-administrative databases We showed that Semantic Web technologies are technically adapted for representing patients' data from medico-administrative databases as RDF and querying them using SPARQL. We also demonstrated that this approach is relevant as it supports the combination of patients' data with hierarchical knowledge in order to address the problem of reconciling precise patients data with more general query criteria. [*O. Dameron, Y. Rivault*] [33], [31], [30]

Telemedicine : ontology-based reasoning and data integration We have developed a system based on a formal ontology that integrates the alert information and the patient data extracted from the electronic health record in order to better classify the importance of alerts. A pilot study was conducted on atrial fibrillation alerts. The results suggest that this approach has the potential to significantly reduce the alert burden in telecardiology. The methods may be extended to other types of connected devices. We also worked on a telemedicine application for monitoring patients with chronic diseases. We proposed an architecture supporting data exchange in the context of multiple chronic diseases [*O. Dameron*] [26], [25], [18]

7.2. Data and knowledge integration based on combinatorial optimization

Participants: Marie Chevallier, Damien Eveillard, Jeanne Got, Julie Laniau, François Moreews, Jacques Nicolas, Anne Siegel.

Packing graphs with ASP for landscape simulation This study is part of a more general research track on graph compression, a fundamental issue for the analysis of biological networks that we address with Answer Set Programming (ASP)modelling. The general issue is to cover a given graph by a set of subgraphs. The IJCAI paper describes an application to crop allocation for generating realistic landscapes. The aim is to cover optimally a bare landscape, represented by its plot graph, with spatial patterns describing local arrangements of crops. This problem belongs to the hard class of graph packing problems. The approach provides a compact solution to the basic problem and at the same time allows extensions such as a flexible integration of expert knowledge. Particular attention is paid to the treatment of symmetries, especially due to sub-graph isomorphism issues. Experiments were conducted on a database of simulated and real landscapes. Currently, our program can process graphs of medium size, a size that enables studies on real agricultural practices. [*J. Nicolas*] [29]

Deciphering transcriptional regulations coordinating the response to environmental changes We introduce a method that extracts from a transcriptional regulatory network determined from a set of predicted transcription factors (TF) and binding site (BS) a subnetwork explaining a given set of observed co-expressions, highlighting those TFs and BSs most likely involved in the co-regulation. The method solves an optimization problem on a graph to select confident paths within the given transcriptional regulatory network joining a putative common regulator with two co-expressed genes via regulatory cascades. It provides a useful modeling scheme for deciphering the regulatory mechanisms that underly the phenotypical response of an organism to environmental challenges and can be used as a reliable tool for further research on genome scale transcriptional regulation studies. [*M. Chevallier, D. Eveillard, A. Siegel*] [13]

Putative bacterial interactions from metagenomic knowledge with an integrative systems ecology approach. Our software tool *shogen* [62] was used to decipher functional roles within a consortium of five mining bacteria through the integration of genomic and metabolic knowledge at genome scale. We first reconstructed a global metabolic network. Next, using a parsimony assumption, we deciphered sets of genes, called Sets from Genome Segments (SGS), that (i) are close on their respective genomes, (ii) take an active part in metabolic pathways and (iii) whose associated metabolic reactions are also closely connected within metabolic networks. The use of SGS (shogen) pinpoints a functional compartmentalization among the investigated species and exhibits putative bacterial interactions necessary for promoting these pathways. [*M. Chevallier, D. Eveillard, A. Siegel*] [15]

Molecular alterations induced by a high-fat high-fiber diet in porcine adipose tissues: variations according to the anatomical fat location Our methods based on the integration of metabolic and regulatory regulations [61] were combined to statistical approaches and applied to the understanding of fatty acid metabolism in porcs and chicken. The analyses evidenced that a high-fat high-fiber diet depressed glucose and lipid anabolic molecular pathways, thus counteracting adipose tissue expansion. Interaction effects between dietary intake of fiber and lipids on gene expression may modulate innate immunity and inflammation, a response which is of interest with regard to chronic inflammation and its adverse effects on health and performance. [*F. Moreews, A. Siegel*] [20]

7.3. Systems biology

Participants: Jérémie Bourdon, Jean Coquet, Victorien Delannée, Jacques Nicolas, Anne Siegel, Nathalie Théret, Pierre Vignet.

Representation of symbolic dynamical systems generated by a substitution. Iterated morphisms are combinatorial processes which are related to several classes of dynamical systems appearing in several fields of computer sciences and mathematics: numeration, ergodic theory, discrete geometry. They may be associated to fractal sets called "Rauzy fractals" whose topological properties are linked to the properties of the underlying dynamical system. We have introduced a generic algorithm framework to check such topological properties within a complete family of iterated morphism. This makes efficient the verification of conjectures on several families of substitutions related to multi-dimensional continued fraction algorithms. [*A. Siegel*] [14]

Identification of logical models for signaling pathways. Logical models of signaling pathways are a promising way of building effective in silico functional models of a cell. The automated learning of Boolean logic models describing signaling pathways can be achieved by training to phosphoproteomics data. In our work, combinatorial optimization methods based on recent logic programming paradigm allow to enumerate, and discriminate the family of logical models explaining data. Together, these approaches enable a robust understanding of the system response. The results are implemented in the *caspo* software. The main weakness of ASP-based learning algorithm is that they focus on the comparison of two time-points and assumes that the system has reached an early steady state. We have generalized such a learning procedure in order to discriminate Boolean networks according to their transient dynamics. To that goal, we exhibit a necessary condition that must be satisfied by a Boolean network dynamics to be consistent with a discretized time series trace. [*A. Siegel*] [23], [28]

Model of the Delayed Translation of Cyclin B Maternal mRNA After Sea Urchin Fertilization. An extended model of the numerical model introduced in [74] was developed to have a better understanding of the role of cyclin B in protein synthesis within minutes after fertilization of sea urchin eggs. The model confirms that regulation of cyclin B biosynthesis is an example of a select protein whose translation is controlled by pathways that are distinct from housekeeping proteins, even though both involve the same cap-dependent initia- tion pathway. Therefore, this model should help provide insight to the signaling utilized for the biosynthesis of cyclin B and other select proteins. [*J. Bourdon, A. Siegel*] [24]

Deciphering pathways involved in TGF-β signalling network. TGF-β is a multifunctional cytokine that regulates mammalian development, differentiation, and homeostasis. As a growth inhibitor of epithelial, endothelial, and hematopoietic cells, TGF-β is a potent anticancer agent in normal tissue. At the opposite TGF-β acts as a promoter of tumor by inducing the hallmarks of the cancer. Consequently targeting the deleterious effects of TGF-β without affecting its physiological role is the common goal of therapeutic strategies. While several strategies based on blocking TGF-β antibodies or small inhibitors of TGF-β receptors have been investigated, they did not take into account the impact of the (extracellular matrix) ECM remodeling that regulates TGF-β bioavailability and the complexity of TGF-β-dependent signaling pathways which regulate both physiological and pathological processes depending on context. In accordance with this, we recently demonstrated the benefical anti-tumor effect of the interplay between TGF-β signaling and the CD103 integrin pathway. At the opposite we have previously demonstrated that the disintegrin ADAMTS1 promotes TGF-β activation in chronic liver disease and we recently characterized interaction with inhibitor peptide to block such effects, using in silico approach. Importantly, we need to take into account a system-wide view and develop predictive models for therapeutic benefit. In that context we demonstrated that the ratio of TGFBR2 to TGFBR1 receptors concentrations can be used to discriminate between metastable regimes of TGF-β signaling model and predic the tumor cell aggressiveness [*N. Théret*][27], [16], [21].

7.4. Sequence and structure annotation

Participants: Aymeric Antoine-Lorquin, Catherine Belleannée, François Coste, Jacques Nicolas.

Detection of mutated primers on metagenomics sequences to detect more species. In targeted metagenomics, an initial task is the detection in each sequence of the primers used for amplifying the targeted region. The selected sequences are then trimmed and clustered in order to inventory species present in the sample. Common pratices consist in retaining only the sequences with perfect primers (i.e. non-mutated by sequencing error). In the context of a study characterizing the biodiversity of tropical soils in unicellular eukaryotes, we have implemented the search for mutated primers, using the grammatical pattern matching tool Logol, and shown that retrieving sequences with mutated primers has a significant impact on targeted metagenomics results, as it makes possible to detect more species (7% additional OTUs in our study) [*A. Antoine-Lorquin, C. Belleannée*] [32], [11].

VIRALpro: a tool to identify viral capsid and tail sequences. Not only sequence data continues to outpace annotation information, but the problem is further exacerbated when organisms are underrepresented in the annotation databases. This is the case with non human-pathogenic viruses which occur frequently in metagenomic projects. Thus there is a need for tools capable of detecting and classifying viral sequences. Based on machine learning techniques, we have proposed a new effective tool for identifying capsid and tail protein sequences, which are the cornerstones toward viral sequence annotation and viral genome classification. The software and corresponding web server are publicly available as part of the SCRATCH suite. [*F. Coste, C. Galiez*] [19]

Learning substitutable context-free grammars to model protein families. In the first experiments on learning substitutable context-free grammars to model protein families, an identified bottleneck for larger scale experimentation was parsing time. We have implemented a new parsing strategy enabling to handle efficiently the ambiguity of 'gap loops', enabling a factor 20 speedup in practice. We have also begun to investigate the inference of more expressive classes, said contextually substitutable, and have proposed a refined graph approach to learn smaller contextually substitutable grammars from smaller training samples in the framework that we have initiated with ReGLiS. [*F. Coste*] [43], [35]

How to measure the topological quality of protein grammars? To assess the quality of grammars modelling protein families, one is interested in their performances to predict new members of the families, classically measured on the basis of recall and precision in the machine learning framework, but also by their modelling power, which is more difficult to evaluate. We propose here to address this later point by measuring the consistency of grammar's parse trees with 3D structures of proteins, when they are available, by the introduction of a set of measures based on respective internal distances. [*F. Coste*] [36]

Tutorial chapter: Learning the Language of Biological Sequences. Learning the language of biological sequences is an appealing challenge for the grammatical inference research field. While some first successes have already been recorded, such as the inference of profile hidden Markov models or stochastic context-free grammars which are now part of the classical bioinformatics toolbox, it is still a source of open and nice inspirational problems for grammatical inference, enabling us to confront our ideas to real fundamental applications. As an introduction to this field, we survey here the main ideas and concepts behind the approaches developed in pattern/motif discovery and grammatical inference to characterize successfully the biological sequences with their specificities. [*F. Coste*] [40]

8. Partnerships and Cooperations

8.1. Regional Initiatives

8.1.1. Regional initiative: the Ecosyst project

Participants: Damien Eveillard, Marie Chevallier, Clémence Frioux, Anne Siegel, Camille Trottier.

EcoSyst is a Biogenouest inter-regional federating project (Brittany â Pays de la Loire) aiming at the emergence of Systems Ecology at the level of western regions. Drawing on the strengths and skills involved, EcoSyst targets the incubation of new ideas and new projects at disciplinary interfaces. Thanks to this community project, we want to develop the skills of Ecology, Environment, Modeling, Bioinformatics and Systems Biology and their application to organisms and ecosystems of interest in agronomy, sea and health. EcoSyst includes also the identification of the major issues and concerns, the fundamental and essential methods and the very real needs of the community (training, tools, ...); this in order to consider the construction of a community platform (or an offer of service within an existing platform) on complex systems modeling, meeting expectations of the community as fully as possible.

8.1.2. Regional partnership with computer science laboratories in Nantes

Participants: Anne Siegel, Jérémie Bourdon, Damien Eveillard, François Coste, Jacques Nicolas.

Methodologies are developed in close collaboration with university of Nantes (LINA) and Ecole centrale Nantes (IRCCyN). This is acted through the Biotempo and Idealg ANR projects and co-development of common software toolboxes within the Renabi-GO platform support. C. Trottier is a co-supervised bioanalysis and software development engineer within the Idealg project. M. Chevallier is a co-supervised development and animation engineer within the regional initiative "Ecosyst". In addition, the former Ph-D student V. Picard and the ongoing Ph-D student J. Laniau are also co-supervised with members of the LINA laboratory.

8.1.3. Regional partnership in Marine Biology

Participants: Catherine Belleannée, Jérémie Bourdon, Jean Coquet, François Coste, Damien Eveillard, Olivier Dameron, Clémence Frioux, Jeanne Got, Julie Laniau, Jacques Nicolas, Camille Trottier, Anne Siegel.

A strong application domain of the Dyliss project is marine Biology. This application domain is co-developed with the station biologique de Roscoff and their three UMR and involves several contracts. The IDEALG consortium is a long term project (10 years, ANR Investissement avenir) aiming at the development of macro-algae biotechnology. Among the research activities, we are particularly interested in the analysis and reconstruction of metabolism and the characterization of key enzymes. Other research contracts concern the modeling of the initiation of sea-urchin translation (former PEPS program Quantoursin, Ligue contre le cancer and ANR Biotempo), the analysis of extremophile archebacteria genomes and their PPI networks (former ANR MODULOME and PhD thesis of P.-F. Pluchon) and the identification of key actors implied in competition for light in the ocean (PELICAN ANR project). In addition, the team participates to a collaboration program (Inria Project Lab) with the Biocore and Ange teams, together with Ifremer-Nantes, focused on the understanding on micro-algae (Ph-D thesis of J. Laniau).

8.1.4. Regional partnership in agriculture and bio-medical domains

Participants: Aymeric Antoine-Lorquin, Catherine Belleannée, François Coste, Jean Coquet, Olivier Dameron, Victorien Delannée, Aurélie Evrard, François Moreews, Jacques Nicolas, Anne Siegel, Nathalie Théret, Denis Tagu, Pierre Vignet.

We have a strong and long term collaboration with biologists of INRA in Rennes : PEGASE and IGEPP units. F. Morrews is a permanent engineer from PEGASE center hosted in the team to develop methods for integrative biology applied to species of interest in agriculture. D. Tagu is a research director at INRA who spends 20% of his time in the team to develop collaborative projects. This partnership is supported by the co-supervision of one post-doctoral student and the co-supervision of several PhD students. The Ph-D thesis of V. Wucher was supported by collaborations with the IGEPP laboratory. The former post-doc of Ch. Bettembourg strengthened these collaborations. This collaboration was also reinforced by collaboration within ANR contracts (MirNadapt, FatInteger). Lately, A. Evrard joined the team at mid-part of her time in collaboration with Agrocampus Ouest and INRA to apply the semantic web to technologies developed within the mirnAdapt framework to new agriculture applications (Brassicae).

We also have a strong and long term collaboration in the bio-medical domain, namely with the IRSET laboratory at Univ. Rennes 1/Irset. N. Théret, research director at INSERM, is hosted in the team to strenghen our collaborative projects. Our collaborations are acted by the co-supervised Ph-D theses of V. Delannée (Metagenotox project, funded by Anses) and J. Coquet. This partnership was reinforced in the former years by the ANR contract Biotempo ended at the end of 2014. In 2015, the project of combining semantic web technologies and bi-clustering classification based on formal concept analysis was applied to systems biology within the PEPS CONFOCAL project. This scientific project has been recently pushed forward in the recent TGFSYSBio project funded by Plan Cancer on the modelling of the microenvironment of TGFbeta signaling network (P. Vignet has been recruited on this contract at the end of 2016).

8.2. National Initiatives

8.2.1. Long-term contracts

8.2.1.1. "Omics"-Line of the Chilean CIRIC-Inria Center

Participants: Meziane Aite, Jérémie Bourdon, François Coste, Marie Chevallier, Damien Eveillard, Clémence Frioux, Jacques Nicolas, Anne Siegel.

We have a cooperation with Univ. of Chile (MATHomics, A. Maass) on methods for the identification of biomarkers and software for biochip design, supported by a national Inria initiative. It aims at combining automatic reasoning on biological sequences and networks with probabilistic approaches to manage, explore and integrate large sets of heterogeneous omics data into networks of interactions allowing to produce biomarkers, with a main application to biomining bacteria. The program is co-funded by Inria and CORFO-chile from 2012 to 2022. In this context, IntegrativeBioChile is an Associate Team between Dyliss and the Laboratory of Bioinformatics and Mathematics of the Genome hosted at Univ. of Chile funded from 2011 to 2016.

8.2.1.2. ANR Idealg

Participants: Jérémie Bourdon, Marie Chevallier, François Coste, Damien Eveillard, Clémence Frioux, Jeanne Got, Jacques Nicolas, Anne Siegel.

IDEALG is one of the five laureates from the national call 2010 for Biotechnology and Bioresource and will run until 2020. It gathers 18 different partners from the academic field (CNRS, IFREMER, UEB, UBO, UBS, ENSCR, University of Nantes, INRA, AgroCampus), the industrial field (C-WEED, Bezhin Rosko, Aleor, France Haliotis, DuPont) as well as a technical center specialized in seaweeds (CEVA) in order to foster biotechnology applications within the seaweed field. It is organized in ten workpackages. We are participating in the tasks related to the establishment of a virtual platform for integrating omics studies on seaweed) and the integrative analysis of seaweed metabolism, in cooperation with SBR Roscoff. Major objectives are the building of brown algae metabolic maps, flux analysis and the selection extraction of important parameters for the production of targeted compounds. We will also contribute to the prediction of specific enzymes (sulfatases) [More details].

8.2.2. Programs funded by research institutions

8.2.2.1. PEPS PEPS: a platform for supporting studies in pharmaco-epidemiology using medico-administrative databases

Participants: Olivier Dameron, Yann Rivault.

As a partner of the PEPS platform, IRISA develops generic methods supporting efficient and semantically-rich queries for pharmaco-epidemiology studies on medico-administrative databases. The leader is Thomas Guyet (IRISA team Lacodam). We showed that Semantic Web technologies are technically suited for representing patients' data from medico-administrative databases as RDF and querying them using SPARQL. We also demonstrated that this approach is relevant as it supports the combination of patients' data with hierarchical knowledge in order to address the problem of reconciling precise patients data with more general query criteria [33], [31], [30]. This work is mostly conducted by Yann Rivault, whose PhD thesis is supervized by Olivier Dameron and Nolwenn LeMeur (Ecole des Hautes Etudes en Santé Publique).

8.2.2.2. Cancer Plan: TGFSYSBIO

Participants: Nathalie Théret, Jacques Nicolas, Olivier Dameron, Anne Siegel, Jean Coquet.

The TGFSYSBIO project aims to develop the first model of extracellular and intracellular TGF-beta system that might permit to analyze the behaviors of TGF-beta activity during the course of liver tumor progression and to identify new biomarkers and potential therapeutic targets. Based on collaboration with Jerome Feret from ENS, Paris,we will combine a rule-based model (Kappa language) to describe extracellular TGF-beta activation and large-scale state-transition based (Cadbiom formalism) model for TGF-beta-dependent intracellular signaling pathways. The multi-scale integrated model will be enriched with a large-scale analysis of liver tissues using shotgun proteomics to characterize protein networks from tumor microenvironment whose remodeling is responsible for extracellular activation of TGF-beta. The trajectories and upstream regulators of the final model will be analyzed with symbolic model checking techniques and abstract interpretation combined with causality analysis. Candidates will be classified with semantic-based approaches and symbolic bi-clustering technics. The project is funded by the national program "Plan Cancer - Systems biology" from 2015 to 2018.

8.2.2.3. ANR Samosa

Participants: Damien Eveillard, Jeanne Got, Anne Siegel.

Oceans are particularly affected by global change, which can cause e.g. increases in average sea temperature and in UV radiation fluxes onto ocean surface or a shrinkage of nutrient-rich areas. This raises the question of the capacity of marine photosynthetic microorganisms to cope with these environmental changes both at short term (physiological plasticity) and long term (e.g. gene alterations or acquisitions causing changes in fitness in a specific niche). Synechococcus cyanobacteria are among the most pertinent biological models to tackle this question, because of their ubiquity and wide abundance in the field, which allows them to be studied at all levels of organization from genes to the global ocean.

The SAMOSA project is funded by ANR from 2014 to 2018, coordinated by F. Gaczarek at the Station Biologique de Roscoff/UPMC/CNRS. The goal of the project is to develop a systems biology approach to characterize and model the main acclimation (i.e., physiological) and adaptation (i.e. evolutionary) mechanisms involved in the differential responses of Synechococcus clades/ecotypes to environmental fluctuations, with the goal to better predict their respective adaptability, and hence dynamics and distribution, in the context of global change. For this purpose, following intensive omics experimental protocol driven by our colleagues from − Station Biologique de Roscoff −, we aim at constructing a gene network model sufficiently flexible to allow the integration of transcriptomic and physiological data.

8.2.2.4. ADT Complex-biomarkers and ADT Proof of concept

Participants: Jeanne Got, Marie Chevallier, Meziane Aite, Anne Siegel.

This project started in Oct. 2014 and aims at designing a working environment based on workflows to assist molecular biologists to integrate large-scale omics data on non-classical species. The main goal of the workflows will be to facilitate the identification of set of regulators involved in the response of a species when challenged by an environmental stress. Applications target extremophile biotechnologies (biomining) and marine biology (micro-algae).

8.2.2.5. ANSES Mecagenotox

Participants: Victorien Delannée, Anne Siegel, Nathalie Théret.

The objective of Mecagenotox project is to characterize and model the human liver ability to bioactivate environmental contaminants during liver chronic diseases in order to assess individual susceptibility. Indeed, liver pathologies which result in the development of fibrosis are associated with a severe dysfunction of liver functions that may lead to increased susceptibility against contaminants. In this project funded by ANSES and coordinated by S. Langouet at IRSET/inserm (Univ. Rennes 1), we will combine cell biology approaches, biochemistry, biophysics, analytical chemistry and bioinformatics to 1) understand how the tension forces induced by the development of liver fibrosis alter the susceptibility of hepatocytes to certain genotoxic chemicals (especially Heterocyclic Aromatic Amines) and 2) model the behavior of xenobiotic metabolism during the liver fibrosis. Our main goal is to identify "sensitive" biomolecules in the network and to understand more comprehensively bioactivation of environmental contaminants involved in the onset of hepatocellular carcinoma.

8.2.2.6. PEPS CONFOCAL

Participants: Olivier Dameron, Jean Coquet, Nathalie Théret, Jacques Nicolas, Anne Siegel, Pierre Vignet.

PEPS CONFOCAL aims at developing new bioinformatics methods for analyzing heterogeneous *omics data and for filtering them according to domain knowledge. The current approaches are facing four main limitations: (1) classic biclustering methods do not support partial overlap of clusters, which is too restrictive considering some genes' pleiotropic nature, (2) they assume that the items to analyze (the genes, the molecules, the signaling pathways...) are independent, (3) they tend to generate numerous clusters leaving to the experts the task of identifying the relevant ones, and (4) they are sensitive to noisy or incomplete data. We investigate the extension of Formal Concept Analysis (FCA) with symbolic knowledge from ontologies in order to process large and complex sets of associations between genes, signaling pathways and the molecules involved in these pathways. Future applications cover the discrete model analysis in molecular biology. CONFOCAL initiated a collaboration with Amedeo Napoli (LORIA Nancy) and Elisabeth Remy (Mathematics Institute Luminy, "Mathematical Methods for Genomics" team).

8.3. European Initiatives

8.3.1. Collaborations with Major European Organizations

Partner: Aachen university (Germany)

Title: Modeling the logical response of a signalling network with constraints-programming.

Partner: Potsdam university (Germany)

Title: Constraint-based programming for the modeling and study of biological networks.

8.4. International Initiatives

8.4.1. Inria International Labs

The Dyliss team is strongly involved in the Inria CIRIC center, and the research line "Omics integrative center". The associated team "IntegrativeBioChile", the post-doc of S. Thiele (2012) and the co-supervision of A. Aravena (2010-2013) contributed to reinforce the complementarity of both Chilean and French teams. In 2013, a workshop was organized in Chile to develop new French-Chilean collaborations within the framework of the CIRIC center. In 2014, Marie Chevallier and Meziane Aite joined the team as engineers to improve softwares resulting from collaborations. Maria-Paz Cortes visited the team during 6 monthes in the framework of her ph-D thesis.

Inria Chile
Associate Team involved in the International Lab:

8.4.1.1. BIOINTEGRATIVECHILE

Title: Integrative Biology in Extreme Environments

International Partner (Institution - Laboratory - Researcher):

Universidad de Chile (Chile) - Center for Mathematical Modeling (CMM) - Maass Alejandro

Start year: 2014

See also: http://www.irisa.fr/dyliss/public/EA/index.html

The project is in the area of bioinformatics, with a special focus on bacteria living in extreme environments, more precisely on microorganisms involved in bio-remediation or bio-production processes. We are particularly interested in bioprocesses such as copper extraction, salmon lethality, metal-resistance, all having an economical interest in Chile. Since the last decade, huge databases of microbial genomic sequences, together with multi-scale and large-scale cellular observations (genomics, transcriptomics, proteomics, metabolomics) have been produced. Each one can be viewed as a different scale of a biological process, either in time or space, but ultimately are related through networks of biological interactions that control the behavior of the system The reconstruction, analysis and modeling of such networks using all levels of information are biologically, mathematically and computationally challenging. Applied on microorganisms living in extreme environments, this question is even more challenging since relatively few knowledge is publicly available on the species, requiring to develop methods which are robust to uncertainty. We are developing methods to integrate and manage heterogeneous omics and uncertain data. This in the purpose of extracting suitable biomarkers from this multi-level information. This question will be addressed by coupling probabilistic and static dynamical systems methods with recent and efficient paradigms of constraint programming (Answer Set Programming).

8.5. International Research Visitors

8.5.1. Visits of International Scientists

- **Argentina**. Fundation Leloir, Buenos Aeres [S. Videla]
- **Chile**. Centro de Modelimiento Matematico, Santiago [A. Maass, N. Loiraâ , M. Latorre, M.-P. Cortes]
- **Niger**. University of Maradi [O. Abdou-Arbi]
- **Germany**. Max Planck Institute for Biophysical Chemistry [C. Galiez]

8.5.2. Research stays abroad

- **Germany**. University of Kaiserslautern [A. Antoine-Lorquin, 2 months]
- **Germany**. University of Potsdam [C. Frioux, 2 months]
- **Japan**. National Institute of Informatics in Tokyo [J.Coquet, 3 months]

8.5.3. Visits to International Teams

- **Chile**. Centro de Modelimiento Matematico, Santiago de Chile [J. Bourdon, M. Aite, F. Coste, A. Siegel]
- **Germany**. Frei Berlin University [A. Siegel]
- **Poland**. Wroclaw University of Science and Technology [F. Coste]
- **Netherland**. Utrecht University [F. Coste]

9. Dissemination

9.1. Promoting Scientific Activities

9.1.1. Scientific Events Selection

9.1.1.1. Member of the Conference Program Committees

- JOBIM (2016): French conference of Bioinformatics [A. Siegel]
- BBCC (2015): Bioinformatica e Biologia Computazionale in Campania [O. Dameron]
- IC 2016 Atelier IA et santé, Symposium sur l'Ingénierie des Connaissances [O. Dameron]
- JFO 2016: Journées Francophones sur les ontologies [O. Dameron]
- SASB 2016: The six international workshop on static analysis and system biology [N.Théret]
- ICGI 2016: The 13th International Conference on Grammatical Inference [F. Coste]

9.1.1.2. Review

- ISMB 2016, Biotechno 2016 [A. Siegel]

9.1.2. Journal

9.1.2.1. Member of the Editorial Boards

- Academic editor: Plos One [J. Bourdon]

9.1.2.2. Reviewer - Reviewing Activities

- Journal of Mathematical Biology. Bioinformatics. Theorie des Sciences Informatiques. [A. Siegel]
- Briefings in Bioinformatics, Journal of Biomedical Informatics, Journal of Biomedical Semantics. [O. Dameron]
- Molecular Cancer, Oncotarget,Hepatology, Int J cancer, Bioinformatics and Biology Insight, Chembioint, Cell death and Disease, Plos One [N. Théret]
- IEEE BIBM 2016: IEEE International Conference on Bioinformatics and Biomedicine [F. Coste]

9.1.3. Invited Talks

- M. Aite *User-control metabolic network reconstruction within flexible workflows with the PADMet Toolbox*, INRA Food Working Group annual assembly, Paris (Jul. 2016)
- F. Coste *Grammatical inference of protein languages*, Seminar of Department of Biomedical Engineering, Wroclaw University of Technology, Polland (May 2016)
- F. Coste *Partial multiple sequence alignments to model protein families*, Theoretical Biology and Bioinformatics Group, Utrecht Univ., Netherland (Sep. 1016)
- F. Coste *Modelling protein families with Protomata-Learner*, University of Chile (Nov. 2016)
- C. Trottier *The PADMET-Toolbox and AuReMe workflow: application to the genome-scale metabolic network reconstruction of algae*, IDEALG project Annual General Meeting, Lorient (Nov. 2016)
- C. Frioux *Metabolic network gap-filling: parsimonious combinatorial methods to approach biological reality*, INRA Food Working Group annual assembly, Paris (Jul. 2016)
- C. Frioux *Answer Set Programming for bioinformatics and metabolic networks*, University of Potsdam Knowledge Processing and Information System group weekly seminar, Potsdam, Germany (Oct. 2016)
- A. Siegel *Cancer biology in the Dyliss Group*, Inria (Mar. 2016)
- A. Siegel *Identification of logical models for signaling pathways: towards a systems biology loop*, EMCTS, Nottingham, UK (Jul. 2016) [39]
- A. Siegel *An introduction to metabolic networks modelling*, INRA (Oct. 2016)
- A. Siegel *A prospective about the construction of bioId chip based on multi-scale integrative methods*, University of Chile (Nov. 2016)
- A. Siegel *Combinatorial problems related to the reconstruction of genome-scale metabolic networks*, University of Lille, Workshop of the BIOSS working group on metabolism (Nov. 2016)
- N.Théret *Computational modeling to identify biomarkers and targets*, DHU2020, Fibrosis and remodelling: from common pathways to personalized targets, Autumn School (Oct. 2016)

9.1.4. Leadership within the Scientific Community

- Member of the steering committee of the International Conference on Grammatical Inference [F. Coste].
- The team was involved in the foundation of a national working group on the symbolic study of dynamical systems named bioss [web access]. The group gathers more 100 scientists, from computer science to biology. Three meetings were organized this year. The group is supported by two French National Research Networks: bioinformatics (GDR BIM : bioinformatique moléculaire) and informatics-mathematics (GDR IM : Informatique Mathématique). It gathered twice in 2016: for a general meeting in Lyon (Jul. 2016) and for a workshop focused on computational and methological insights about metabolic network in Lille (Nov. 2016) [A. Siegel]

9.1.5. Scientific Expertise

- Member (nominated) CNU section 65 [O. Dameron]
- Scientific Advisory Board of GDR BIM " Molecular Bioinformatics" [J. Nicolas].
- Inria National evaluation board [A. Siegel]
- Member of the Operational Legal and Ethical Risk Assessment Committee (COERLE) at Inria [J. Nicolas].
- Recruitment committees: Professor (UMPC, Paris) [A. Siegel, N Théret], Associate professor (Nancy) [O. Dameron), Engineer (INRA) [A. Siegel], Inria senior researcher (National committee) [A. Siegel].
- Member of the IRISA laboratory council [F. Coste].
- Member of the Inria Rennes center council [A. Siegel].

- Scientific Advisory Board of Biogenouest [J. Bourdon, N Théret].
- Member of SCAS (Service Commun d'Action Sociale) of Univ. Rennes 1 [C. Belleannée].
- Member of CUMIR (Commission des Utilisateurs des Moyens Informatiques, Inria Rennes) [F. Coste].
- Expertise for Prix Victor et Erminia MESCLE [N. Théret]
- Expertise for Ligue Contre le Cancer, InterRégion Rhône-Alpes-Auvergne-Drome-Saone et Loire [N. Théret]
- Member of the board of directors of the French Society for biology of the extracellular matrix [N. Théret]

9.2. Teaching - Supervision - Juries

9.2.1. Teaching track responsibilities

- Coordination of the doctoral school "Life, Agronomy and Health" of University of Rennes 1 [N. Théret]
- Coordination of the master degree "Bioinformatics and genomics", Univ. Rennes1 [O. Dameron]
- Coordination of the track "From Data to Knowledge: Machine Learning, Modeling and Indexing Multimedia Contents and Symbolic Data", Master in Computer Science, University of Rennes 1, France [F. Coste].

9.2.2. Course responsibilities

- "Bioinformatique expérimentale", Master 1 in computer science, Univ. Rennes1 & ENS [O. Dameron]
- "Bases de mathématiques et probablité" and "Méthodes en informatique", Master1 in public health, Univ. Rennes 1 [O. Dameron]
- "Représentation des connaissances biomédicales", Master 2 in public health, Univ. Rennes 1 [O. Dameron]
- "Principes de programmation et d'algorithmique", Master 1 in bioinformatics, Univ. Rennes 1 [O. Dameron]
- "Gestion de projets informatiques", Master 1 in bioinformatics, Univ. Rennes 1 [O. Dameron]
- "Standardisation des connaissances et bio-ontologies", Master 2 in bioinformatics, Univ. Rennes 1 [O. Dameron]
- "e-Santé et réseaux hospitaliers", last year in engineering school ESIR, Univ. Rennes 1, [O. Dameron]
- "Equilibre Dynamique de la communication Cellulaire" Master 2 in Sciences cellulaire et Moléculaire du Vivant, Univ. Rennes 1 [N. Theret]

9.2.3. Teaching

Licence: C. Belleannée, Langages formels, 22h, L3 informatique, Univ. Rennes1, France.

Licence: C. Belleannée, Traitement de textes et données tabulées, 40h, L1 informatique, Univ. Rennes1, France.

Licence: C. Belleannée, Algorithmique et Programmation Fonctionnelle, 60, L1 informatique, Univ. Rennes1, France.

Licence: J. Coquet, Algorithmique et Programmation Fonctionnelle, 40h, L1 informatique, Rennes1, France.

Licence: O. Dameron, Biostatistiques, 12h, PACES, Univ. Rennes 1, France.

Licence: V. Delannée, Bureautique, 36h, DFGSP2, Univ. Rennes 1, France.

Licence: C. Frioux, Bureautique, 12h, L1 informatique, Rennes1, France.

Master: C. Belleannée, Algorithmique du texte et bioinformatique, 10h, M1 informatique, Univ. Rennes1, France

Master: C. Belleannée, Préférences, Logique et Contraintes, 40h M1 informatique,Univ. Rennes1, France

Master: F. Coste, Apprentissage Supervisé, 10h, M2 Informatique, Univ. Rennes 1, France

Master: F. Coste, Données Séquentielles Symboliques, 10h, M2 Informatique, Univ. Rennes 1, France

Master: O. Dameron, Bases de mathématiques et probablité, 30h, Master1 in public health, Univ. Rennes 1, France.

Master: O. Dameron, Méthodes en informatique, 50h, Master1 in public health, Univ. Rennes 1, France.

Master: O. Dameron, Bioinformatique expérimentale, 10h, M1 informatique, Univ. Rennes 1 and ENS Rennes, France.

Master: O. Dameron, Principes de programmation et algorithmique, 50h, M1 bioinformatique et génomique, Univ. Rennes 1, France.

Master: O. Dameron, Gestion de projets informatiques, 23h, M1 bioinformatique et génomique, Univ. Rennes 1, France.

Master: O. Dameron, Standardisation des connaissances et bio-ontologies, 24h, M2 bioinformatique et génomique, Univ. Rennes 1, France.

Master: O. Dameron, Représentation des connaissances biomédicales, 20h, M2 bioinformatique et génomique, Univ. Rennes 1, France.

Master: A. Siegel, Integrative and Systems biology, 20h, M2, Univ. Rennes 1, France

Master: N. Théret, Extracellular matrix remodeling and Signaling, 3H, Univ. Rennes 1, France

Master: N. Théret, Extracellular matrix remodeling and Signaling, 3H, Univ. Cergy Pontoise, France

Doctorat: A. Siegel, Modelling the integration of heterogeneous knowledge with Answer Set Programming, 4h, Ecole de printemps, Porquerolles, France

9.2.4. Supervision

HDR: Olivier Dameron *Ontology-based methods for analyzing life science data* [12]

PhD : Aymeric Antoine-Lorquin, *TITRE*, started in Oct. 2013, supervised by C. Belleannée, defended on the 1st of December 2016 [11]

PhD in progress : Lucas Bourneuf, *Justifiable graph decomposition to assist biological network understanding*, started in Oct. 2016, supervised by J. Nicolas.

PhD in progress : Jean Coquet, *Semantic-based reasoning for biological pathways analysis*, started in Oct. 2014, supervised by O. Dameron and N. Théret.

PhD in progress : Victorien Delannée, *Optimisation à différentes échelles pour étudier la variabilité de la toxicité de contaminants alimentaires*, started in Oct. 2014, supervised by A. Siegel and N. Théret.

PhD in progress : Clémence Frioux, *Using preferences in Answer Set Programming to decipher interactions within the species of an ecosystem at the genomic scale*, started in Oct. 2015, supervised by A. Siegel.

PhD in progress : Julie Laniau, *Méthodes d'optimisation combinatoire pour reconstruire et analyser les systèmes métaboliques de microalgues*, started in Oct. 2013, supervised by A. Siegel and D. Eveillard.

PhD in progress : Yann Rivault, *Analyse de parcours de soins à partir de bases de données médico-administratives en utilisant des outils du Web Sémantique : identification de complications et de leurs déterminants suite à la pose chirurgicale de dispositif médical implantable en ambulatoire* , started in Oct. 2015, supervised by O. Dameron and N. Lemeur.

9.2.5. Juries

- *Member of Ph-D thesis juries*. M. Morterol, Univ. Paris Sud [A. Siegel, reviewer]. A. Rougny, Univ. Paris Sud [A. Siegel, reviewer]. P. Traynard, ENS Paris [A. Siegel, jury member]. A Lamora, Univ. Nantes [N. Théret, reviewer]. L. Alcaraz, Univ. Lyon [N. Théret, reviewer], F Courivaud, UMPC [N. Théret, reviewer]. P Hascoet, Univ. Rennes1 [N. Théret, president]
- *Member of habilitation thesis juries*. O. Dameron, Univ. Rennes 1 [A. Siegel, jury member], A. Chateau, Univ. Montpellier [A. Siegel, reviewer]. M. Elati, Univ. Evry [A. Siegel, reviewer]. L. Levy, Univ Paris-Diderot [N. Théret, reviewer]. C. Le Goff, Univ Paris Descartes [N. Théret, reviewer].
- *Member of medical thesis jury*. P. Hamon, Rennes [O. Dameron, jury member].

9.2.6. Internships

- Internship, from Jan until Jun 2016. Supervised by A. Siegel. Student: Mael Conan. Subject: Reconstruction of the metabolic map of E. Synecchococcus.
- Internship, from Jan until Jun 2016. Supervised by M. Chevallier and A. Siegel. Student: Pierre Vignet. Subject: Development of a web interface for the aided-curation of metabolic network identifiers.
- Internship, from Jun until Jul 2016. Supervised by O. Dameron and A. Siegel. Student: David Saulpic. Subject: Using formal concept analysis to classify the attractors of perturbated boolean networks.
- Internship, from Jan. until Jun 2016. Supervised by J. Nicolas. Student: Lucas Bourneuf. Subject: Model reduction with power graph algorithms.
- Internship, from Feb. until Jun 2016. Supervised by F. Coste. Student: Mikael Demirdelen. Subject: Fast parser for biological sequences and a new algorithm for the inference of substitutable languages.
- Internship, from Jun until Jul 2016. Supervised by O. Dameron. Student: Arnaud Belcour. Subject: Intégration de données biologiques en RDF pour l'analyse de réseaux de régulation.
- Internship, from Jun until Jul 2016. Supervised by O. Dameron. Student: Mael Kerbiriou. Subject: Création et analyse d'un réseau de régulation génique en RDF : application au puceron.
- Internship, from Jun until Jul 2016. Supervised by A. Evrard. Student: Xavier Garnier. Subject: Mise à jour et développement d'AskOmics, outil d'intégration et d'interrogation de données biologiques.
- Internship, from Jan. until Jun. 2016. Supervised by C. Belleannée. Student: Nathan Alary. Subject: Données génomiques et données ChIP-Seq au service de la prédiction de sites de fixation d'un facteur de transcription. Application au facteur LXRalpha.
- Internship, from Mar until Aug 2016. Supervised by J. Got. Student: Sanae El Mhijar, Subject: Analyse et vérifications du réseau métabolique de *Tisochrysis lutea*.
- Internship, from May until Jul 2016. Supervised by F. Coste. Student: RemySun, Subject: Learning Deep Latent Features of Proteins.
- Internship, from Jun. until Jul. 2016. Supervised by F. Morreews. Student: Vivien Le Breton Subject: RDF et SPARQL pour l'integration de réseaux métaboliques et génétiques de référence.
- Internship, from Jan. until Jun. 2016. Supervised by J. Nicolas. Student: Guillaume Lebreton Subject: Metabolic pathway reconstruction on metagenomes, application to the development of a bacterial consortium for fermented products.
- Internship, from Apr. until Jun. 2016. Supervised by J. Nicolas. Student: Marie Salmon Subject: Analyse par concepts formels de données génomiques sur le mélanome du chien.

9.3. Popularization

- *Organization of Sciences en Cour[t]s*. Since 2007, Sciences en Cour[t]s is a project of Nicomaque organization, the association of PhD and PhD students of Brittany. It is a popularization Festival where PhD students explain their thesis via short films of 5min. The goal is to present their scientific researches to the general public. Every year, PhD students of Inria/IRISA join the organization or make movies.[*J. Coquet (coordinator of the festival), V. Delannée (president of Nicomaque), A. Antoine-Lorquin (organizer of the festival)*] [more info].
- *Production of Sciences en Cour[t]s film*. "Une petite histoire de symbiose(s)" .[*C. Frioux*] [more info].
- *Bioinfo-fr.net* Bioinfo-fr.net is a french web site where researchers, engineers and students talks about bioinformatics. We have written or contributed to 3 articles for this web site on diverse subjects: "Remise des diplômes du master BIG (Rennes)", "Les dev' jam c'est bon pour vous !", "Ecrire son parseur à la main: chroniques d'une mauvaise bonne idée". [*L Bourneuf, O. Dameron*]. [more info].

10. Bibliography

Major publications by the team in recent years

[1] C. Belleannée, O. Sallou, J. Nicolas.*Logol: Expressive Pattern Matching in sequences. Application to Ribosomal Frameshift Modeling*, in "PRIB2014 - Pattern Recognition in Bioinformatics, 9th IAPR International Conference", Stockholm, Sweden, M. Comin, L. Kall, E. Marchiori, A. Ngom, J. Rajapakse (editors), Springer International Publishing, August 2014, vol. 8626, p. 34-47 [*DOI :* 10.1007/978-3-319-09192-1_4], https://hal.inria.fr/hal-01059506.

[2] J. Bourdon, D. Eveillard, A. Siegel.*Integrating quantitative knowledge into a qualitative gene regulatory network*, in "PLoS Computational Biology", September 2011, vol. 7, n^o^ 9 [*DOI :* 10.1371/journal.pcbi.1002157], http://hal.archives-ouvertes.fr/hal-00626708.

[3] A. Bretaudeau, F. Coste, F. Humily, L. Garczarek, G. Le Corguille, C. Six, M. Ratin, O. Collin, W. M. Schluchter, F. Partensky.*CyanoLyase: a database of phycobilin lyase sequences, motifs and functions*, in "Nucleic Acids Research", November 2012, vol. 41 [*DOI :* 10.1093/nar/gks1091], http://hal.inria.fr/hal-00760946.

[4] F. Coste, G. Kerbellec.*A Similar Fragments Merging Approach to Learn Automata on Proteins*, in "ECML:Machine Learning: ECML 2005, 16th European Conference on Machine Learning, Porto, Portugal, October 3-7, 2005, Proceedings", J. Gama, R. Camacho, P. Brazdil, A. Jorge, L. Torgo (editors), Lecture Notes in Computer Science, Springer, 2005, vol. 3720, p. 522-529.

[5] M. Gebser, C. Guziolowski, M. Ivanchev, T. Schaub, A. Siegel, P. Veber, S. Thiele.*Repair and Prediction (under Inconsistency) in Large Biological Networks with Answer Set Programming*, in "Principles of Knowledge Representation and Reasoning", AAAI Press, 2010.

[6] C. Guziolowski, A. Bourdé, F. Moreews, A. Siegel.*BioQuali Cytoscape plugin: analysing the global consistency of regulatory networks*, in "Bmc Genomics", 2009, vol. 26, n^o^ 10, 244 [*DOI :* 10.1186/1471-2164-10-244], http://hal.inria.fr/inria-00429804.

[7] C. Guziolowski, S. Videla, F. Eduati, S. Thiele, T. Cokelaer, A. Siegel, J. Saez-Rodriguez.*Exhaustively characterizing feasible logic models of a signaling network using Answer Set Programming*, in "Bioinformatics", August 2013, vol. 29, n^o 18, p. 2320-2326 [*DOI :* 10.1093/bioinformatics/btt393], http://hal.inria.fr/hal-00853704.

[8] J. Nicolas, P. Durand, G. Ranchy, S. Tempel, A.-S. Valin.*Suffix-Tree Analyser (STAN): looking for nucleotidic and peptidic patterns in genomes*, in "Bioinformatics (Oxford, England)", 2005, vol. 21, p. 4408-4410, http://hal.archives-ouvertes.fr/hal-00015234.

[9] S. Prigent, G. Collet, S. M. Dittami, L. Delage, F. Ethis de Corny, O. Dameron, D. Eveillard, S. Thiele, J. Cambefort, C. Boyen, A. Siegel, T. Tonon.*The genome-scale metabolic network of Ectocarpus siliculosus (EctoGEM): a resource to study brown algal physiology and beyond*, in "Plant Journal", September 2014, p. 367-81 [*DOI :* 10.1111/tpj.12627], https://hal.archives-ouvertes.fr/hal-01057153.

[10] C. Rousseau, M. Gonnet, M. Le Romancer, J. Nicolas.*CRISPI: a CRISPR interactive database*, in "Bioinformatics", 2009, vol. 25, n° 24, p. 3317-3318.

Publications of the year

Doctoral Dissertations and Habilitation Theses

[11] A. Antoine-Lorquin, A. Siegel, C. Belleannée.*Interest of grammatical modelling to achieve pattern matching in genomic sequences*, Université de Rennes 1, December 2016, https://hal.inria.fr/tel-01416734.

[12] O. Dameron.*Ontology-based methods for analyzing life science data*, Univ. Rennes 1, January 2016, Habilitation à diriger des recherches, https://hal.inria.fr/tel-01403371.

Articles in International Peer-Reviewed Journal

[13] V. Acuña, A. Aravena, C. Guziolowski, D. Eveillard, A. Siegel, A. Maass.*Deciphering transcriptional regulations coordinating the response to environmental changes*, in "BMC Bioinformatics", January 2016, vol. 17, n° 1, p. 129-42, https://hal.archives-ouvertes.fr/hal-01260866.

[14] V. Berthé, J. Bourdon, T. Jolivet, A. Siegel.*A combinatorial approach to products of Pisot substitutions*, in "Ergodic Theory and Dynamical Systems", 2016, vol. 36, n° 6, p. 1757-1794 [*DOI :* 10.1017/etds.2014.141], https://hal.inria.fr/hal-01196326.

[15] P. Bordron, M. Latorre, M.-P. Cortés, M. González, S. Thiele, A. Siegel, A. Maass, D. Eveillard.*Putative bacterial interactions from metagenomic knowledge with an integrative systems ecology approach*, in "MicrobiologyOpen", January 2016, vol. 5, n° 1, p. 106-17, https://hal.archives-ouvertes.fr/hal-01289230.

[16] M. Boutet, L. Gauthier, M. Leclerc, G. Gros, V. De Montpreville, N. Théret, E. Donnadieu, F. Mami-Chouaib.*TGF-β signaling intersects with CD103 integrin signaling to promote T lymphocyte accumulation and antitumor activity in the lung tumor microenvironment*, in "Cancer Research", 2016, vol. 76, n° 7, p. 1757-69 [*DOI :* 10.1158/0008-5472.CAN-15-1545], https://hal-univ-rennes1.archives-ouvertes.fr/hal-01282442.

[17] F. J. Cousin, S. Jouan-Lanhouet, N. Théret, C. Brenner, E. Jouan, G. Le Moigne-Muller, M.-T. Dimanche-Boitrel, G. Jan.*The probiotic Propionibacterium freudenreichii*

as a new adjuvant for TRAIL-based therapy in colorectal cancer, in "Oncotarget", January 2016 [*DOI :* 10.18632/ONCOTARGET.6881], https://hal-univ-rennes1.archives-ouvertes.fr/hal-01259215.

[18] P. FINET, B. GIBAUD, O. DAMERON, R. LEBOUQUIN JEANNES.*Relevance of health level 7 clinical document architecture and integrating the healthcare enterprise cross-enterprise document sharing profile for managing chronic wounds in a telemedicine context*, in " Healthcare Technology Letter ", 2016, vol. 3, n^o 1, p. 22-26 [*DOI :* 10.1049/HTL.2015.0053], https://hal.archives-ouvertes.fr/hal-01301007.

[19] C. GALIEZ, M. CHRISTOPHE, F. COSTE, P. BALDI.*VIRALpro: a tool to identify viral capsid and tail sequences*, in "Bioinformatics", 2016 [*DOI :* 10.1093/BIOINFORMATICS/BTV727], https://hal.archives-ouvertes.fr/hal-01242251.

[20] F. GONDRET, A. VINCENT, M. HOUÉE-BIGOT, A. SIEGEL, S. LAGARRIGUE, I. LOUVEAU, D. CAUSEUR.*Molecular alterations induced by a high-fat high-fiber diet in porcine adipose tissues: variations according to the anatomical fat location*, in "BMC Genomics", December 2016, vol. 17, n^o 1, article number 120 [*DOI :* 10.1186/S12864-016-2438-3], https://hal.archives-ouvertes.fr/hal-01286555.

[21] M.-A. LAURENT, D. BONNIER, N. THÉRET, P. TUFFÉRY, G. MOROY.*In silico characterization of the interaction between LSKL peptide, a LAP-TGF-beta derived peptide, and ADAMTS1*, in "Computational Biology and Chemistry", April 2016, p. 155-161 [*DOI :* 10.1016/J.COMPBIOLCHEM.2016.01.012], https://hal.inria.fr/hal-01401685.

[22] G. NICOLAS, D. WALLON, C. CHARBONNIER, O. QUENEZ, S. ROUSSEAU, A.-C. RICHARD, A. ROVELET-LECRUX, S. COUTANT, K. LE GUENNEC, D. BACQ, J.-G. GARNIER, R. OLASO, A. BOLAND, V. MEYER, J.-F. DELEUZE, H. M. MUNTER, G. BOURQUE, D. AULD, A. MONTPETIT, M. LATHROP, L. GUYANT-MARÉCHAL, O. MARTINAUD, J. PARIENTE, A. ROLLIN-SILLAIRE, F. PASQUIER, I. LE BER, M. SARAZIN, B. CROISILE, C. BOUTOLEAU-BRETONNIÈRE, C. THOMAS-ANTÉRION, C. PAQUET, M. SAUVEE, O. MOREAUD, A. GABELLE, F. SELLAL, M. CECCALDI, L. CHAMARD, F. BLANC, T. FREBOURG, D. CAMPION, D. HANNEQUIN.*Screening of dementia genes by whole-exome sequencing in early-onset Alzheimer disease: input and lessons*, in "European Journal of Human Genetics", May 2016, vol. 24, n^o 5, p. 710-716 [*DOI :* 10.1038/EJHG.2015.173], https://hal.archives-ouvertes.fr/hal-01431285.

[23] M. OSTROWSKI, L. PAULEVÉ, T. SCHAUB, A. SIEGEL, C. GUZIOLOWSKI.*Boolean Network Identification from Perturbation Time Series Data combining Dynamics Abstraction and Logic Programming*, in "BioSystems", 2016 [*DOI :* 10.1016/J.BIOSYSTEMS.2016.07.009], https://hal.archives-ouvertes.fr/hal-01354075.

[24] V. PICARD, O. MULNER-LORILLON, J. BOURDON, J. MORALES, P. CORMIER, A. SIEGEL, R. BELLÉ.*Model of the delayed translation of cyclin B maternal mRNA after sea urchin fertilization*, in "Molecular Reproduction and Development", 2016 [*DOI :* 10.1002/MRD.22746], http://hal.upmc.fr/hal-01390047.

[25] A. ROSIER, P. MABO, L. TEMAL, P. VAN HILLE, O. DAMERON, L. DELÉGER, C. GROUIN, P. ZWEIGENBAUM, J. JACQUES, E. CHAZARD, L. LAPORTE, C. HENRY, A. BURGUN.*Personalized and automated remote monitoring of atrial fibrillation*, in "EP-Europace", 2016, vol. 18, n^o 3, p. 347-352 [*DOI :* 10.1093/EUROPACE/EUV234], https://hal-univ-rennes1.archives-ouvertes.fr/hal-01331019.

[26] A. ROSIER, P. MABO, L. TEMAL, P. VAN HILLE, O. DAMERON, L. DELÉGER, C. GROUIN, P. ZWEIGENBAUM, J. JACQUES, E. CHAZARD, L. LAPORTE, C. HENRY, A. BURGUN.*Remote Monitoring of Cardiac Implantable Devices: Ontology Driven Classification of the Alerts*, in "Studies in Health Technology and In-

formatics", 2016, vol. 221, p. 59–63 [*DOI :* 10.3233/978-1-61499-633-0-59], https://hal-univ-rennes1.archives-ouvertes.fr/hal-01319949.

[27] S. SWARUP SAMAL, A. NALDI, D. GRIGORIEV, A. WEBER, N. THÉRET, O. RADULESCU.*Geometric analysis of pathways dynamics: application to versatility of TGF-β receptors*, in "BioSystems", July 2016, vol. 149, p. 3–14 [*DOI :* 10.1016/J.BIOSYSTEMS.2016.07.004], https://hal.inria.fr/hal-01379033.

[28] S. VIDELA, J. SAEZ-RODRIGUEZ, C. GUZIOLOWSKI, A. SIEGEL.*caspo: a toolbox for automated reasoning on the response of logical signaling networks families*, in "Bioinformatics", 2016 [*DOI :* 10.1093/BIOINFORMATICS/BTW738], https://hal.inria.fr/hal-01426880.

International Conferences with Proceedings

[29] T. GUYET, Y. MOINARD, J. NICOLAS, R. QUINIOU.*Packing graphs with ASP for landscape simulation*, in "IJCAI 2016 - 25th International joint conference on artificial intelligence", New-york, United States, AAAI, July 2016, 8, https://hal.inria.fr/hal-01327368.

National Conferences with Proceeding

[30] Y. RIVAULT, N. LE MEUR, O. DAMERON.*Complications post-opératoires et mode de prise en charge en angioplastie : apport du Programme de Médicalisation des Systèmes d'Information (PMSI)*, in "Congrès Adelf-Epiter 2016", Rennes, France, September 2016, https://hal.inria.fr/hal-01389688.

[31] Y. RIVAULT, N. LE MEUR, O. DAMERON.*La gestion de données médico-administratives grâce aux outils du Web sémantique*, in "Journées ADELF-EMOIS « Système d'information hospitalier et Epidémiologie »", Dijon, France, Revue d'Épidémiologie et de Santé Publique, March 2016, vol. 64, n^{o} Supplement 1, S15 [*DOI :* 10.1016/J.RESPE.2016.01.051], https://hal.archives-ouvertes.fr/hal-01297389.

Conferences without Proceedings

[32] A. ANTOINE-LORQUIN, F. MAHÉ, M. DUNTHORN, C. BELLEANNÉE.*Impact de la recherche d'amorces mutées sur les résultats d'analyses métagénomiques*, in "Journées Ouvertes en Biologie, Informatique et Mathématiques (JOBIM)", Rennes, France, Jobim-2016, Société Française de Bioinformatique (SFBI), June 2016, https://hal.archives-ouvertes.fr/hal-01343121.

[33] F. BALUSSON, M.-A. BOTREL, O. DAMERON, Y. DAUXAIS, E. DREZEN, A. DUPUY, T. GUYET, D. GROSS-AMBLARD, A. HAPPE, N. LE MEUR, B. LE NAUTOUT, E. LERAY, E. NOWAK, C. RAULT, E. OGER, E. POLARD.*PEPS: a platform for supporting studies in pharmaco-epidemiology using medico-administrative databases*, in "International Congress on e-Health Research", Paris, France, October 2016, https://hal.inria.fr/hal-01380939.

[34] M. CHEVALLIER, M. AITE, J. GOT, G. COLLET, N. LOIRA, M.-P. CORTÉS, C. FRIOUX, J. LANIAU, C. TROTTIER, A. MAASS, A. SIEGEL.*Handling the heterogeneity of genomic and metabolic networks data within flexible workflows with the PADMet toolbox*, in "Jobim 2016 : 17ème Journées Ouvertes en Biologie, Informatique et Mathématiques", Lyon, France, June 2016, https://hal.inria.fr/hal-01377844.

[35] F. COSTE, M. DEMIRDELEN.*A refined parsing graph approach to learn smaller contextually substitutable grammars with less data*, in "ICGI 2016 - 13th International Conference on Grammatical Inference", Delft, Netherlands, October 2016, https://hal.inria.fr/hal-01406337.

[36] W. DYRKA, F. COSTE, O. UNOLD, L. CULER, A. KACZMAREK.*How to measure the topological quality of protein grammars?*, in "ICGI 2016 - 13th International Conference on Grammatical Inference", Delft, Netherlands, October 2016, https://hal.inria.fr/hal-01406331.

[37] F. LEGEAI, C. BETTEMBOURG, A. BRETAUDEAU, Y. CHAUSSIN, O. DAMERON, D. TAGU.*BIPAA/Askomics, a new and easy approach for querying genomics and epigenomics elements in interaction*, in "XXVth International Congress of Entomology 2016,", Orlando, Florida, United States, September 2016, https://hal.inria.fr/hal-01391080.

[38] F. MOREEWS, O. SALLOU, O. COLLIN. *An application suite based on the IFB Container as a Service platform*, in "ECCB 2016", Den Haag, Netherlands, ISCB, September 2016, https://hal.inria.fr/hal-01394295.

[39] A. SIEGEL.*A system biology loop for the identification of boolean networks to model signaling networks.* , in "European Conference on Mathematical and Theoretical Biology", Nottingham, France, 2016, https://hal.inria.fr/hal-01399456.

Scientific Books (or Scientific Book chapters)

[40] F. COSTE.*Learning the Language of Biological Sequences*, in "Topics in Grammatical Inference", J. HEINZ, J. M. SEMPERE (editors), Springer-Verlag, 2016 [*DOI :* 10.1007/978-3-662-48395-4_8], https://hal.inria.fr/hal-01244770.

Other Publications

[41] F. COUSIN, S. JOUAN-LANHOUET, N. THÉRET, C. BRENNER, E. JOUAN, G. LE MOIGNE-MULLER, M.-T. DIMANCHE-BOITREL.*The probiotic [i]Propionibacteriumfreudenreichii[/i] as a new adjuvant for TRAIL-based therapy in colorectal cancer*, September 2016, 4 . International Symposium on Propionibacteria and Bibfidobacteria , Poster, https://hal.archives-ouvertes.fr/hal-01372131.

[42] F. COUSIN, S. JOUAN-LANHOUET, N. THÉRET, C. BRENNER, E. JOUAN, G. LE MOIGNE-MULLER, M.-T. DIMANCHE-BOITREL, G. JAN.*The probiotic [i]Propionibacterium freudenreichii[/i] as a new adjuvant for TRAIL-based therapy in colorectal cancer*, June 2016, International Scientific Conference on Probiotics and Prebiotics - IPC2016, Poster, https://hal.archives-ouvertes.fr/hal-01286497.

[43] M. DEMIRDELEN.*Fast parser for biological sequences and a new algorithm for the inference of substitutable languages*, Univ. Rennes 1, June 2016, https://hal.inria.fr/hal-01406352.

[44] A. EVRARD, C. BETTEMBOURG, O. DAMERON, O. FILANGI, A. BRETAUDEAU, F. F. LEGEAI, R. DELOURME, M. MANZANARES-DAULEUX, M. JUBAULT.*Comparative genomic analysis of Clubroot resistance in the Brassicaceae family*, October 2016, Brassica2016, Poster, https://hal.inria.fr/hal-01391707.

[45] A. EVRARD, C. BETTEMBOURG, M. M. JUBAULT, O. DAMERON, O. FILANGI, A. BRETAUDEAU, F. F. LEGEAI.*Integration and query of biological datasets with Semantic Web technologies: AskOmics*, June 2016, Journées Ouvertes Biologie, Informatique et Mathématiques (JOBIM 2016), Poster, https://hal.inria.fr/hal-01391087.

[46] S. RÉMY.*Apprentissage profond et acquisition de représentations latentes de séquences peptidiques*, ENS Rennes, August 2016, https://hal.inria.fr/hal-01406368.

References in notes

[47] O. ABDOU-ARBI, S. LEMOSQUET, J. VAN MILGEN, A. SIEGEL, J. BOURDON.*Exploring metabolism flexibility in complex organisms through quantitative study of precursor sets for system outputs*, in "BMC Systems Biology", 2014, vol. 8, n° 1, 8 [*DOI :* 10.1186/1752-0509-8-8], https://hal.inria.fr/hal-00947219.

[48] V. ACUÑA, A. ARAVENA, A. MAASS, A. SIEGEL.*Modeling parsimonious putative regulatory networks: complexity and heuristic approach*, in "15th conference in Verification, Model Checking, and Abstract Interpretation", San Diego, United States, Springer, 2014, vol. 8318, p. 322-336 [*DOI :* 10.1007/978-3-642-54013-4_18], https://hal.inria.fr/hal-00926477.

[49] G. ANDRIEUX, M. LE BORGNE, N. THÉRET.*An integrative modeling framework reveals plasticity of TGF-Beta signaling*, in "BMC Systems Biology", 2014, vol. 8, n° 1, 30 [*DOI :* 10.1186/1752-0509-8-30], http://www.hal.inserm.fr/inserm-00978313.

[50] C. BARAL.*Knowledge Representation, Reasoning and Declarative Problem Solving*, Cambridge University Press, 2010.

[51] T. BAUMURATOVA, D. SURDEZ, B. DELYON, G. STOLL, O. DELATTRE, O. RADULESCU, A. SIEGEL.*Localizing potentially active post-transcriptional regulations in the Ewing's sarcoma gene regulatory network*, in "BMC Systems Biology", 2010, vol. 4, n° 1, 146 [*DOI :* 10.1186/1752-0509-4-146], http://www.hal.inserm.fr/inserm-00984711.

[52] R. BELLAZZI.*Big Data and Biomedical Informatics: A Challenging Opportunity*, in "Yearbook of medical informatics", 2014, vol. 9, n° 1, In press.

[53] R. BELLAZZI, M. DIOMIDOUS, I. N. SARKAR, K. TAKABAYASHI, A. ZIEGLER, A. T. MCCRAY.*Data analysis and data mining: current issues in biomedical informatics*, in "Methods of information in medicine", 2011, vol. 50, n° 6, p. 536–544.

[54] C. BELLEANNÉE, O. SALLOU, J. NICOLAS.*Logol: Expressive Pattern Matching in sequences. Application to Ribosomal Frameshift Modeling*, in "PRIB2014 - Pattern Recognition in Bioinformatics, 9th IAPR International Conference", Stockholm, Sweden, M. COMIN, L. KALL, E. MARCHIORI, A. NGOM, J. RAJAPAKSE (editors), Springer International Publishing, August 2014, vol. 8626, p. 34-47 [*DOI :* 10.1007/978-3-319-09192-1_4], https://hal.inria.fr/hal-01059506.

[55] R. BELLÉ, S. PRIGENT, A. SIEGEL, P. CORMIER.*Model of cap-dependent translation initiation in sea urchin: a step towards the eukaryotic translation regulation network*, in "Molecular Reproduction and Development", 2010, vol. 77, n° 3, p. 257-64.

[56] R. BENDAOUD, A. NAPOLI, Y. TOUSSAINT.*Formal Concept Analysis: A Unified Framework for Building and Refining Ontologies*, in "Knowledge Engineering: Practice and Patterns, 16th International Conference, EKAW 2008, Acitrezza, Italy, September 29 - October 2, 2008. Proceedings", A. GANGEMI, J. EUZENAT (editors), Lecture Notes in Computer Science, Springer, 2008, vol. 5268, p. 156-171, http://dx.doi.org/10.1007/978-3-540-87696-0_16.

[57] C. BETTEMBOURG, C. DIOT, O. DAMERON.*Optimal threshold determination for interpreting semantic similarity and particularity*, in "PLoS ONE", 2015, vol. 10, n° 7, e0133579 [*DOI :* 10.1371/JOURNAL.PONE.0133579], https://hal.archives-ouvertes.fr/hal-01207763.

[58] C. BETTEMBOURG, C. DIOT, O. DAMERON.*Optimal Threshold Determination for Interpreting Semantic Similarity and Particularity: Application to the Comparison of Gene Sets and Metabolic Pathways Using GO and ChEBI*, in "PLoS ONE", 2015, 30 [*DOI :* 10.1371/JOURNAL.PONE.0133579], https://hal.inria.fr/hal-01184934.

[59] C. BIZER, T. HEATH, T. BERNERS LEE.*Linked Data–The story so far*, in "International Journal on Semantic Web and Information Systems", 2009, vol. 5, n^o 3, p. 1–22.

[60] J. A. BLAKE, C. J. BULT.*Beyond the data deluge: Data integration and bio-ontologies*, in "Journal of Biomedical Informatics", 2006, vol. 39, n^o 3, p. 314–320.

[61] P. BLAVY, F. GONDRET, S. LAGARRIGUE, J. VAN MILGEN, A. SIEGEL.*Using a large-scale knowledge database on reactions and regulations to propose key upstream regulators of various sets of molecules participating in cell metabolism*, in "BMC Systems Biology", 2014, vol. 8, n^o 1, 32 [*DOI :* 10.1186/1752-0509-8-32], https://hal.inria.fr/hal-00980499.

[62] P. BORDRON, D. EVEILLARD, A. MAASS, A. SIEGEL.*An ASP application in integrative biology: identification of functional gene units*, in "LPNMR - 12th Conference on Logic Programming and Nonmonotonic Reasoning - 2013", Corunna, Spain, September 2013, http://hal.inria.fr/hal-00853762.

[63] H. CHEN, T. YU, J. Y. CHEN.*Semantic Web meets Integrative Biology: a survey*, in "Briefings in bioinformatics", 2012, vol. 14, n^o 1, p. 109–125.

[64] G. COLLET, D. EVEILLARD, M. GEBSER, S. PRIGENT, T. SCHAUB, A. SIEGEL, S. THIELE.*Extending the Metabolic Network of Ectocarpus Siliculosus using Answer Set Programming*, in "LPNMR - 12th Conference on Logic Programming and Nonmonotonic Reasoning - 2013", Corunna, Spain, September 2013, http://hal.inria.fr/hal-00853752.

[65] F. COSTE, G. GARET, A. GROISILLIER, J. NICOLAS, T. TONON.*Automated Enzyme classification by Formal Concept Analysis*, in "ICFCA - 12th International Conference on Formal Concept Analysis", Cluj-Napoca, Romania, Springer, June 2014, https://hal.inria.fr/hal-01063727.

[66] S. DAMINELLI, V. J. HAUPT, M. REIMANN, M. SCHROEDER.*Drug repositioning through incomplete bicliques in an integrated drug-target-disease network*, in "Integr. Biol.", 2012, vol. 4, p. 778-788, http://dx.doi.org/10.1039/C2IB00154C.

[67] O. DEMEURE, F. LECERF, C. DUBY, C. DESERT, S. DUCHEIX, H. GUILLOU, S. LAGARRIGUE.*Regulation of LPCAT3 by LXR*, in "Gene", Jan 2011, vol. 470, n^o 1-2, p. 7–11.

[68] S. M. DITTAMI, T. BARBEYRON, C. BOYEN, J. CAMBEFORT, G. COLLET, L. DELAGE, A. GOBET, A. GROISILLIER, C. LEBLANC, G. MICHEL, D. SCORNET, A. SIEGEL, J. E. TAPIA, T. TONON.*Genome and metabolic network of "Candidatus Phaeomarinobacter ectocarpi" Ec32, a new candidate genus of Alphaproteobacteria frequently associated with brown algae*, in "Frontiers in Genetics", 2014, vol. 5, 241 [*DOI :* 10.3389/FGENE.2014.00241], https://hal.inria.fr/hal-01079739.

[69] M. GEBSER, R. KAMINSKI, B. KAUFMANN, T. SCHAUB.*Answer Set Solving in Practice*, Synthesis Lectures on Artificial Intelligence and Machine Learning, Morgan and Claypool Publishers, 2012.

[70] F. GONDRET, I. LOUVEAU, M. HOUEE, D. CAUSEUR, A. SIEGEL.*Data integration*, in "Meeting INRA-ISU", Ames, United States, March 2015, 11, https://hal.archives-ouvertes.fr/hal-01210940.

[71] F. GONDRET, A. VINCENT, M. HOUEE, S. LAGARRIGUE, A. SIEGEL, D. CAUSEUR, I. LOUVEAU.*Integrative responses of pig adipose tissues to high-fat high-fiber diet: towards key regulators of energy flexibility*, March 2015, ASAS/ADSA midwest meeting, Poster, https://hal.archives-ouvertes.fr/hal-01210925.

[72] F. HERAULT, A. VINCENT, O. DAMERON, P. LE ROY, P. CHEREL, M. DAMON.*The longissimus and semimembranosus muscles display marked differences in their gene expression profiles in pig*, in "PLoS ONE", 2014, vol. 9, n° 5, e96491 [*DOI :* 10.1371/JOURNAL.PONE.0096491], https://hal.inria.fr/hal-00989635.

[73] R. KAMINSKI, A. SIEGEL, T. SCHAUB, S. VIDELA.*Minimal Intervention Strategies in Logical Signaling Networks with ASP*, in "Theory and Practice of Logic Programming", September 2013, vol. 13, n° Special issue 4-5, p. 675-690 [*DOI :* 10.1017/S1471068413000422], http://hal.inria.fr/hal-00853747.

[74] S. LAURENT, A. RICHARD, O. MULNER-LORILLON, J. MORALES, D. FLAMENT, V. GLIPPA, J. BOURDON, P. GOSSELIN, A. SIEGEL, P. CORMIER, R. BELLÉ.*Modelization of the regulation of protein synthesis following fertilization in sea urchin shows requirement of two processes: a destabilization of eIF4E:4E-BP complex and a great stimulation of the 4E-BP-degradation mechanism, both rapamycin-sensitive*, in "Frontiers in Genetics", 2014, vol. 5, 117 [*DOI :* 10.3389/FGENE.2014.00117], https://hal.inria.fr/hal-01079758.

[75] F. LEGEAI, T. DERRIEN, V. WUCHER, D. AUDREY, G. LE TRIONNAIRE, D. TAGU.*Long non-coding RNA in the pea aphid; identification and comparative expression in sexual and asexual embryos*, in "Arthropod Genomics Symposium", Urbana, United States, June 2014, https://hal.inria.fr/hal-01091304.

[76] S.-W. LEUNG, C. MELLISH, D. ROBERTSON.*Basic Gene Grammars and DNA-ChartParser for language processing of Escherichia coli promoter DNA sequences*, in "Bioinformatics", 2001, vol. 17, n° 3, p. 226-236 [*DOI :* 10.1093/BIOINFORMATICS/17.3.226], http://bioinformatics.oxfordjournals.org/content/17/3/226.abstract.

[77] A. LIHONOSOVA, A. KAMINSKAYA.*Using Formal Concept Analysis for Finding the Closest Relatives among a Group of Organisms*, in "Procedia Computer Science", 2014, vol. 31, n° Complete, p. 860-868, http://dx.doi.org/10.1016/j.procs.2014.05.337.

[78] K. M. LIVINGSTON, M. BADA, W. A. BAUMGARTNER, L. E. HUNTER.*KaBOB: ontology-based semantic integration of biomedical databases*, in "BMC bioinformatics", 2015, vol. 16, 126.

[79] N. MESSAI, M.-D. DEVIGNES, A. NAPOLI, M. SMAIL-TABBONE.*Many-Valued Concept Lattices for Conceptual Clustering and Information Retrieval*, in "Proceedings of the 2008 Conference on ECAI 2008: 18th European Conference on Artificial Intelligence", Amsterdam, The Netherlands, The Netherlands, IOS Press, 2008, p. 127–131, http://dl.acm.org/citation.cfm?id=1567281.1567313.

[80] M. OBITKO, V. SNÁSEL, J. SMID.*Ontology Design with Formal Concept Analysis*, in "Proceedings of the CLA 2004 International Workshop on Concept Lattices and their Applications, Ostrava, Czech Republic, September 23-24, 2004", V. SNÁSEL, R. BELOHLÁVEK (editors), CEUR Workshop Proceedings, CEUR-WS.org, 2004, vol. 110, http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-110/paper12.pdf.

[81] P.-F. PLUCHON, T. FOUQUEAU, C. CREZE, S. LAURENT, J. BRIFFOTAUX, G. HOGREL, A. PALUD, G. HENNEKE, A. GODFROY, W. HAUSNER, M. THOMM, J. NICOLAS, D. FLAMENT.*An Extended Network of Genomic Maintenance in the Archaeon Pyrococcus abyssi Highlights Unexpected Associations between Eucaryotic Homologs*, in "PLoS ONE", 2013, vol. 8, n^o^ 11, e79707 [*DOI :* 10.1371/JOURNAL.PONE.0079707], http://hal.inria.fr/hal-00911795.

[82] S. PRIGENT, G. COLLET, S. M. DITTAMI, L. DELAGE, F. ETHIS DE CORNY, O. DAMERON, D. EVEILLARD, S. THIELE, J. CAMBEFORT, C. BOYEN, A. SIEGEL, T. TONON.*The genome-scale metabolic network of Ectocarpus siliculosus (EctoGEM): a resource to study brown algal physiology and beyond*, in "Plant Journal", September 2014, p. 367-81 [*DOI :* 10.1111/TPJ.12627], https://hal.archives-ouvertes.fr/hal-01057153.

[83] M. ROUANE-HACENE, M. HUCHARD, A. NAPOLI, P. VALTCHEV.*Using Formal Concept Analysis for Discovering Knowledge Patterns*, in "CLA'10: 7th International Conference on Concept Lattices and Their Applications", Sevilla, Spain, S. O. MARZENA KRYSZKIEWICZ (editor), CEUR, University of Sevilla, October 2010, n^o^ 672, p. 223-234, http://hal-lirmm.ccsd.cnrs.fr/lirmm-00531802.

[84] C. ROUSSEAU, M. GONNET, M. LE ROMANCER, J. NICOLAS.*CRISPI: a CRISPR interactive database*, in "Bioinformatics", 2009, vol. 25, n^o^ 24, p. 3317-3318.

[85] L. ROYER, M. REIMANN, B. ANDREOPOULOS, M. SCHROEDER.*Unraveling Protein Networks with Power Graph Analysis*, in "PLoS Comput Biol", 07 2008, vol. 4, n^o^ 7, e1000108, http://dx.plos.org/10.1371%2Fjournal.pcbi.1000108.

[86] T. SCHAUB, A. SIEGEL, S. VIDELA.*Reasoning on the response of logical signaling networks with Answer Set Programming*, in "Logical Modeling of Biological Systems", Wiley Online Librairy, 2014, p. 49-92 [*DOI :* 10.1002/9781119005223.CH2], https://hal.inria.fr/hal-01079762.

[87] T. SCHAUB, S. THIELE.*Metabolic Network Expansion with Answer Set Programming*, in "ICLP 2009", LNCS, Springer, 2009, vol. 5649, p. 312-326.

[88] S. SCHULZ, L. BALKANYI, R. CORNET, O. BODENREIDER.*From Concept Representations to Ontologies: A Paradigm Shift in Health Informatics?*, in "Healthcare informatics research", 2013, vol. 19, n^o^ 4, p. 235-242.

[89] D. SEARLS.*The language of genes*, in "Nature", 2002, vol. 420, p. 211-217.

[90] D. SEARLS.*String Variable Grammar: A Logic Grammar Formalism for the Biological Language of DNA*, in "Journal of Logic Programming", 1995, vol. 24, n^o^ 1&2, p. 73-102.

[91] Z. D. STEPHENS, S. Y. LEE, F. FAGHRI, R. H. CAMPBELL, C. ZHAI, M. J. EFRON, R. IYER, M. C. SCHATZ, S. SINHA, G. E. ROBINSON.*Big Data: Astronomical or Genomical?*, in "PLoS biology", 2015, vol. 13, n^o^ 7, e1002195.

[92] S. TEMPEL, C. ROUSSEAU, F. TAHI, J. NICOLAS.*ModuleOrganizer: detecting modules in families of transposable elements*, in "BMC Bioinformatics", 2010, vol. 11, 474 [*DOI :* 10.1186/1471-2105-11-474], http://hal.inria.fr/inria-00536742.

[93] S. THIELE, J. SAEZ-RODRIGUEZ, L. CERONE, A. SIEGEL, C. GUZIOLOWSKI, S. KLAMT.*Extended notions of sign consistency to relate experimental data to signaling and regulatory network topologies*, in

"BMC Bioinformatics", 2015, vol. 16, 345 [*DOI :* 10.1186/s12859-015-0733-7], https://hal.inria.fr/hal-01225228.

[94] S. VIDELA, C. GUZIOLOWSKI, F. EDUATI, S. THIELE, M. GEBSER, J. NICOLAS, J. SAEZ-RODRIGUEZ, T. SCHAUB, A. SIEGEL.*Learning Boolean logic models of signaling networks with ASP*, in "Journal of Theoretical Computer Science (TCS)", September 2015, n^o 599, p. 79-101 [*DOI :* 10.1016/j.tcs.2014.06.022], https://hal.inria.fr/hal-01058610.

[95] S. VIDELA, I. KONOKOTINA, L. ALEXOPOULOS, J. SAEZ-RODRIGUEZ, T. SCHAUB, A. SIEGEL, C. GUZIOLOWSKI.*Designing experiments to discriminate families of logic models*, in "Frontiers in Bioengineering and Biotechnology", September 2015, 9 [*DOI :* 10.3389/fbioe.2015.00131], https://hal.inria.fr/hal-01196178.

[96] R. WILLE.*restructuring lattice theory: an approach based on hierarchies of concepts*, in "Proceedings of the 7th International Conference on Formal Concept Analysis", Berlin, Heidelberg, ICFCA '09, Springer-Verlag, 2009, p. 314–339, http://dx.doi.org/10.1007/978-3-642-01815-2_23.

[97] V. WUCHER, D. TAGU, J. NICOLAS. , B. LAUSEN, S. KROLAK-SCHWERDT, M. BÖHMER (editors)*Edge Selection in a Noisy Graph by Concept Analysis – Application to a Genomic Network*, Data Science, Learning by Latent Structures, and Knowledge Discovery, Springer, 2014, 550, https://hal.inria.fr/hal-01093337.

[98] V. WUCHER. *Modeling of a gene network between mRNAs and miRNAs to predict gene functions involved in phenotypic plasticity in the pea aphid*, Universite Rennes 1, November 2014, https://hal.archives-ouvertes.fr/tel-01095967.

Project-Team FLUMINANCE

Fluid Flow Analysis, Description and Control from Image Sequences

IN COLLABORATION WITH: Institut de recherche mathématique de Rennes (IRMAR)

IN PARTNERSHIP WITH:
Institut national de recherche en sciences et technologies pour l'environnement et l'agriculture

RESEARCH CENTER
Rennes - Bretagne-Atlantique

THEME
Earth, Environmental and Energy Sciences

Table of contents

Project-Team FLUMINANCE

Creation of the Project-Team: 2009 July 01

Keywords:

Computer Science and Digital Science:

3. - Data and knowledge
3.3. - Data and knowledge analysis
3.4. - Machine learning and statistics
5.3. - Image processing and analysis
5.4. - Computer vision
5.9. - Signal processing
6. - Modeling, simulation and control
6.1. - Mathematical Modeling
6.2. - Scientific Computing, Numerical Analysis & Optimization
6.3. - Computation-data interaction
6.4. - Automatic control

Other Research Topics and Application Domains:

3.2. - Climate and meteorology
3.3. - Geosciences
5. - Industry of the future
5.2. - Design and manufacturing

1. Members

Research Scientists

Etienne Mémin [Team leader, Inria, Senior Researcher, HDR]
Jocelyne Erhel [Inria, Senior Researcher, HDR]
Christophe Collewet [IRSTEA, Researcher, HDR]
Dominique Heitz [IRSTEA, Researcher]
Cédric Herzet [Inria, Researcher]

Faculty Member

Roger Lewandowski [Univ. Rennes I, Professor]

Technical Staff

Yvan Crenner [Inria]
Lionel Lenôtre [Inria, until Jun 2016]

PhD Students

Pranav Chandramouli [Inria]
Benjamin Delfino [Inria, until Oct 2016]
Bastien Hamlat [IFPEN, from Oct 2016]
Musaab Khalid Osman Mohammed [IRSTEA]
Benoit Pinier [Univ. Rennes I]
Valentin Resseguier [Inria, granted by IFREMER]
Romain Schuster [ITGA]

Post-Doctoral Fellows

Pierre Derian [Inria]
Mamadou Diallo [Inria, from Oct 2016]
David Imberti [Inria, until Jul 2016]
Thibaut Tronchin [Inria, from May 2016]

Visiting Scientists
Shengze Cai [Zhejiang University, from Nov 2016]
Gisela Daniela Charo [University of Buenos Aires, from Sep 2016 until Oct 2016]

Administrative Assistant
Huguette Béchu [Inria]

Others
Jean-Raynald de Dreuzy [CNRS, Research Scientist, HDR]
Johan Carlier [Irstea, Research Engineer]

2. Overall Objectives

2.1. Overall Objectives

The research group that we have entitled FLUMINANCE from a contraction between the words "Fluid" and "Luminance" is dedicated to the extraction of information on fluid flows from image sequences and to the development of tools for the analysis and control of these flows. The objectives of the group are at the frontiers of several important domains that range from fluid mechanics to geophysics. One of the main originality of the FLUMINANCE group is to combine cutting-edge researches on data-assimilation and flow numerical modeling with an ability to conduct proper intensive experimental validations on prototype flows mastered in laboratory. The scientific objectives decompose in four main themes:

- **Fluid flows characterization from images**
  In this first axis, we aim at providing accurate measurements and consistent analysis of complex fluid flows through image analysis techniques.The application domain ranges from industrial processes and experimental fluid mechanics to environmental sciences. This theme includes also the use of non-conventional imaging techniques such as Schlieren techniques, Shadowgraphs, holography. The objective will be here to go towards 3D dense velocity measurements.
- **Coupling dynamical model and image data**
  We focus here on the study, through image data, of complex and partially known fluid flows involving complex boundary conditions, multi-phase fluids, fluids and structures interaction problems. Our credo is that image analysis can provide sufficiently fine observations on small and medium scales to construct models which, applied at medium and large scale, account accurately for a wider range of the dynamics scales. The image data and a sound modeling of the dynamical uncertainty at the observation scale should allow us to reconstruct the observed flow and to provide efficient real flows (experimental or natural) based dynamical modeling. Our final goal will be to go towards a 3D reconstruction of real flows, or to operate large motion scales simulations that fit real world flow data and incorporate an appropriate uncertainty modeling.
- **Control and optimization of turbulent flows**

We are interested on active control and more precisely on closed-loop control. The main idea is to extract reliable image features to act on the flow. This approach is well known in the robot control community, it is called visual servoing. More generally, it is a technique to control a dynamic system from image features. We plan to apply this approach on flows involved in various domains such as environment, transport, microfluidic, industrial chemistry, pharmacy, food industry, agriculture, etc.

- **Numerical models for geophysical flows simulation and analysis** Numerical models are very useful for environmental applications. Several difficulties must be handled simultaneously, in a multidisciplinary context. For example, in geophysics, media are highly heterogeneous and only few data are available. Stochastic models are often necessary to describe unresolved physical processes. Computational domains are characterized by complex 3D geometries, requiring adapted space discretization. Equations modeling flow and transport are transient, requiring also adapted time discretization. Moreover, these equations can be coupled together or with other equations in a global nonlinear system. These large-scale models are very time and memory consuming. High performance computing is thus required to run these types of scientific simulations. Supercomputers and clusters are quite powerful, provided that the numerical models are written with a parallel paradigm.

3. Research Program

3.1. Estimation of fluid characteristic features from images

The measurement of fluid representative features such as vector fields, potential functions or vorticity maps, enables physicists to have better understanding of experimental or geophysical fluid flows. Such measurements date back to one century and more but became an intensive subject of research since the emergence of correlation techniques [47] to track fluid movements in pairs of images of a particles laden fluid or by the way of clouds photometric pattern identification in meteorological images. In computer vision, the estimation of the projection of the apparent motion of a 3D scene onto the image plane, referred to in the literature as optical-flow, is an intensive subject of researches since the 80's and the seminal work of B. Horn and B. Schunk [57]. Unlike to dense optical flow estimators, the former approach provides techniques that supply only sparse velocity fields. These methods have demonstrated to be robust and to provide accurate measurements for flows seeded with particles. These restrictions and their inherent discrete local nature limit too much their use and prevent any evolutions of these techniques towards the devising of methods supplying physically consistent results and small scale velocity measurements. It does not authorize also the use of scalar images exploited in numerous situations to visualize flows (image showing the diffusion of a scalar such as dye, pollutant, light index refraction, flurocein,...). At the opposite, variational techniques enable in a well-established mathematical framework to estimate spatially continuous velocity fields, which should allow more properly to go towards the measurement of smaller motion scales. As these methods are defined through PDE's systems they allow quite naturally constraints to be included such as kinematic properties or dynamic laws governing the observed fluid flows. Besides, within this framework it is also much easier to define characteristic features estimation procedures on the basis of physically grounded data model that describes the relation linking the observed luminance function and some state variables of the observed flow. The Fluminance group has allowed a substantial progress in this direction with the design of dedicated dense estimation techniques to estimate dense fluid motion fields. See [8] for a detailed review. More recently problems related to scale measurement and uncertainty estimation have been investigated [51]. Dynamically consistent and highly robust techniques have been also proposed for the recovery of surface oceanic streams from satellite images [49].

3.2. Data assimilation and Tracking of characteristic fluid features

Real flows have an extent of complexity, even in carefully controlled experimental conditions, which prevents any set of sensors from providing enough information to describe them completely. Even with the highest levels of accuracy, space-time coverage and grid refinement, there will always remain at least a lack of resolution and some missing input about the actual boundary conditions. This is obviously true for the complex flows encountered in industrial and natural conditions, but remains also an obstacle even for standard academic flows thoroughly investigated in research conditions.

This unavoidable deficiency of the experimental techniques is nevertheless more and more compensated by numerical simulations. The parallel advances in sensors, acquisition, treatment and computer efficiency allow the mixing of experimental and simulated data produced at compatible scales in space and time. The inclusion of dynamical models as constraints of the data analysis process brings a guaranty of coherency based on fundamental equations known to correctly represent the dynamics of the flow (e.g. Navier Stokes equations) [11]. Conversely, the injection of experimental data into simulations ensures some fitting of the model with reality.

To enable data and models coupling to achieve its potential, some difficulties have to be tackled. It is in particular important to outline the fact that the coupling of dynamical models and image data are far from being straightforward. The first difficulty is related to the space of the physical model. As a matter of fact, physical models describe generally the phenomenon evolution in a 3D Cartesian space whereas images provides generally only 2D tomographic views or projections of the 3D space on the 2D image plane. Furthermore, these views are sometimes incomplete because of partial occlusions and the relations between the model state variables and the image intensity function are otherwise often intricate and only partially known. Besides, the dynamical model and the image data may be related to spatio-temporal scale spaces of very different natures which increases the complexity of an eventual multiscale coupling. As a consequence of these difficulties, it is necessary generally to define simpler dynamical models in order to assimilate image data. This redefinition can be done for instance on an uncertainty analysis basis, through physical considerations or by the way of data based empirical specifications. Such modeling comes to define inexact evolution laws and leads to the handling of stochastic dynamical models. The necessity to make use and define sound approximate models, the dimension of the state variables of interest and the complex relations linking the state variables and the intensity function, together with the potential applications described earlier constitute very stimulating issues for the design of efficient data-model coupling techniques based on image sequences.

On top of the problems mentioned above, the models exploited in assimilation techniques often suffer from some uncertainties on the parameters which define them. Hence, a new emerging field of research focuses on the characterization of the set of achievable solutions as a function of these uncertainties. This sort of characterization indeed turns out to be crucial for the relevant analysis of any simulation outputs or the correct interpretation of operational forecasting schemes. In this context, the tools provided by the Bayesian theory play a crucial role since they encompass a variety of methodologies to model and process uncertainty. As a consequence, the Bayesian paradigm has already been present in many contributions of the Fluminance group in the last years and will remain a cornerstone of the new methodologies investigated by the team in the domain of uncertainty characterization.

This wide theme of research problems is a central topic in our research group. As a matter of fact, such a coupling may rely on adequate instantaneous motion descriptors extracted with the help of the techniques studied in the first research axis of the FLUMINANCE group. In the same time, this coupling is also essential with respect to visual flow control studies explored in the third theme. The coupling between a dynamics and data, designated in the literature as a Data Assimilation issue, can be either conducted with optimal control techniques [58], [59] or through stochastic filtering approaches [52], [55]. These two frameworks have their own advantages and deficiencies. We rely indifferently on both approaches.

3.3. Optimization and control of fluid flows with visual servoing

Fluid flow control is a recent and active research domain. A significant part of the work carried out so far in that field has been dedicated to the control of the transition from laminarity to turbulence. Delaying, accelerating or modifying this transition is of great economical interest for industrial applications. For instance, it has been shown that for an aircraft, a drag reduction can be obtained while enhancing the lift, leading consequently to limit fuel consumption. In contrast, in other application domains such as industrial chemistry, turbulence phenomena are encouraged to improve heat exchange, increase the mixing of chemical components and enhance chemical reactions. Similarly, in military and civilians applications where combustion is involved, the control of mixing by means of turbulence handling rouses a great interest, for example to limit infra-red signatures of fighter aircraft.

Flow control can be achieved in two different ways: passive or active control. Passive control provides a permanent action on a system. Most often it consists in optimizing shapes or in choosing suitable surfacing (see for example [50] where longitudinal riblets are used to reduce the drag caused by turbulence). The main problem with such an approach is that the control is, of course, inoperative when the system changes. Conversely, in active control the action is time varying and adapted to the current system's state. This approach requires an external energy to act on the system through actuators enabling a forcing on the flow through for instance blowing and suction actions [62], [54]. A closed-loop problem can be formulated as an optimal control issue where a control law minimizing an objective cost function (minimization of the drag, minimization of the actuators power, etc.) must be applied to the actuators [48]. Most of the works of the literature indeed comes back to open-loop control approaches [61], [56], [60] or to forcing approaches [53] with control laws acting without any feedback information on the flow actual state. In order for these methods to be operative, the model used to derive the control law must describe as accurately as possible the flow and all the eventual perturbations of the surrounding environment, which is very unlikely in real situations. In addition, as such approaches rely on a perfect model, a high computational costs is usually required. This inescapable pitfall has motivated a strong interest on model reduction. Their key advantage being that they can be specified empirically from the data and represent quite accurately, with only few modes, complex flows' dynamics. This motivates an important research axis in the Fluminance group.

3.4. Numerical models applied to hydrogeology and geophysics

The team is strongly involved in numerical models for hydrogeology and geophysics. There are many scientific challenges in the area of groundwater simulations. This interdisciplinary research is very fruitful with cross-fertilizing subjects. For example, high performance simulations were very helpful for finding out the asymptotic behaviour of the plume of solute transported by advection-dispersion. Numerical models are necessary to understand flow transfer in fractured media.

The team develops stochastic models for groundware simulations as well as for oceanic and atmospheric flows. Numerical models must then include Uncertainty Quantification methods, spatial and time discretization. Then, the discrete problems must be solved with efficient algorithms. The team develops parallel algorithms for complex numerical simulations and conducts performance analysis.

3.5. Numerical algorithms and high performance computing

Linear algebra is at the kernel of most scientific applications, in particular in physical or chemical engineering. For example, steady-state flow simulations in porous media are discretized in space and lead to a large sparse linear system. The target size is 10^7 in 2D and 10^{10} in 3D. For transient models such as diffusion, the objective is to solve about 10^4 linear systems for each simulation. Memory requirements are of the order of Giga-bytes in 2D and Tera-bytes in 3D. CPU times are of the order of several hours to several days. Several methods and solvers exist for large sparse linear systems. They can be divided into three classes: direct, iterative or semi-iterative. Direct methods are highly efficient but require a large memory space and a rapidly increasing computational time. Iterative methods of Krylov type require less memory but need a scalable preconditioner to remain competitive. Iterative methods of multigrid type are efficient and scalable, used by themselves or as preconditioners, with a linear complexity for elliptic or parabolic problems but they are not so efficient for hyperbolic problems. Semi-iterative methods such as subdomain methods are hybrid direct/iterative methods which can be good tradeoffs. The convergence of iterative and semi-iterative methods and the accuracy of the results depend on the condition number which can blow up at large scale. The objectives are to analyze the complexity of these different methods, to accelerate convergence of iterative methods, to measure and improve the efficiency on parallel architectures, to define criteria of choice.

In geophysics, a main concern is to solve inverse problems in order to fit the measured data with the model. Generally, this amounts to solve a linear or nonlinear least-squares problem. Complex models are in general coupled multi-physics models. For example, reactive transport couples advection-diffusion with chemistry. Here, the mathematical model is a set of nonlinear Partial Differential Algebraic Equations. At each timestep of an implicit scheme, a large nonlinear system of equations arise. The challenge is to solve efficiently and accurately these large nonlinear systems.

Approximation in Krylov subspace is in the core of the team activity since it provides efficient iterative solvers for linear systems and eigenvalue problems as well. The later are encountered in many fields and they include the singular value problem which is especially useful when solving ill posed inverse problems.

4. Application Domains

4.1. Introduction

By designing new approaches for the analysis of fluid-image sequences the FLUMINANCE group aims at contributing to several application domains of great interest for the community and in which the analysis of complex fluid flows plays a central role. The group focuses mainly on two broad application domains:

- Environmental sciences;
- Experimental fluid mechanics and industrial flows.

We detail hereafter these two application domains.

4.2. Environmental sciences

The first huge application domain concerns all the sciences that aim at observing the biosphere evolution such as meteorology, climatology or oceanography but also remote sensing study for the monitoring of meteorological events or human activities consequences. For all these domains image analysis is a practical and unique tool to *observe, detect, measure, characterize or analyze* the evolution of physical parameters over a large domain. The design of generic image processing techniques for all these domains might offer practical software tools to measure precisely the evolution of fluid flows for weather forecasting or climatology studies. It might also offer possibilities of close surveillance of human and natural activities in sensible areas such as forests, river edges, and valley in order to monitor pollution, floods or fire. The need in terms of local weather forecasting, risk prevention, or local climate change is becoming crucial for our tomorrow's life. At a more local scale, image sensors may also be of major utility to analyze precisely the effect of air curtains for safe packaging in agro-industrial.

4.3. Experimental fluid mechanics and industrial flows

In the domain of **experimental fluid mechanics**, the visualization of fluid flows plays a major role, especially for turbulence study since high frequency imaging has been made currently available. Together with analysis of turbulence at different scales, one of the major goals pursued at the moment by many scientists and engineers consists in studying the ability to manipulate a flow to induce a desired change. This is of huge technological importance to enhance or inhibit mixing in shear flows, improve energetic efficiency or control the physical effects of strain and stresses. This is for instance of particular interest for:

- military applications, for example to limit the infra-red signatures of fighter aircraft;
- aeronautics and transportation, to limit fuel consumption by controlling drag and lift effects of turbulence and boundary layer behavior;
- industrial applications, for example to monitor flowing, melting, mixing or swelling of processed materials, or preserve manufactured products from contamination by airborne pollutants, or in industrial chemistry to increase chemical reactions by acting on turbulence phenomena.

.

5. New Software and Platforms

5.1. 2DLayeredMotion

Estimation of 2D independent mesoscale layered atmospheric motion fields
FUNCTIONAL DESCRIPTION

This software enables to estimate a stack of 2D horizontal wind fields corresponding to a mesoscale dynamics of atmospheric pressure layers. This estimator is formulated as the minimization of a global energy function. It relies on a vertical decomposition of the atmosphere into pressure layers. This estimator uses pressure data and classification clouds maps and top of clouds pressure maps (or infra-red images). All these images are routinely supplied by the EUMETSAT consortium which handles the Meteosat and MSG satellite data distribution. The energy function relies on a data model built from the integration of the mass conservation on each layer. The estimator also includes a simplified and filtered shallow water dynamical model as temporal smoother and second-order div-curl spatial regularizer. The estimator may also incorporate correlation-based vector fields as additional observations. These correlation vectors are also routinely provided by the Eumetsat consortium.

- Participant: Etienne Memin
- Contact: Etienne Memin
- URL: http://fluid.irisa.fr/index.html

5.2. 3DLayeredMotion

Estimation of 3D interconnected layered atmospheric motion fields
FUNCTIONAL DESCRIPTION

This software extends the previous 2D version. It allows (for the first time to our knowledge) the recovery of 3D wind fields from satellite image sequences. As with the previous techniques, the atmosphere is decomposed into a stack of pressure layers. The estimation relies also on pressure data and classification clouds maps and top of clouds pressure maps. In order to recover the 3D missing velocity information, physical knowledge on 3D mass exchanges between layers has been introduced in the data model. The corresponding data model appears to be a generalization of the previous data model constructed from a vertical integration of the continuity equation.

- Contact: Etienne Memin
- URL: http://fluid.irisa.fr

5.3. DenseMotion

Estimation of 2D dense motion fields
FUNCTIONAL DESCRIPTION

This code allows the computation from two consecutive images of a dense motion field. The estimator is expressed as a global energy function minimization. The code enables the choice of different data models and different regularization functionals depending on the targeted application. Generic motion estimators for video sequences or fluid flows dedicated estimators can be set up. This software allows in addition the users to specify additional correlation based matching measurements. It enables also the inclusion of a temporal smoothing prior relying on a velocity vorticity formulation of the Navier-Stoke equation for Fluid motion analysis applications.

- Participant: Etienne Memin
- Contact: Etienne Memin
- URL: http://fluid.irisa.fr/index.html

5.4. Low-Order-Motion

Estimation of low order representation of fluid motion
FUNCTIONAL DESCRIPTION

This code enables the estimation of a low order representation of a fluid motion field from two consecutive images.The fluid motion representation is obtained using a discretization of the vorticity and divergence maps through regularized Dirac measure. The irrotational and solenoidal components of the motion fields are expressed as linear combinations of basis functions obtained through the Biot-Savart law. The coefficient values and the basis function parameters are formalized as the minimizer of a functional relying on an intensity variation model obtained from an integrated version of the mass conservation principle of fluid mechanics.

- Participants: Etienne Memin and Anne Cuzol
- Contact: Etienne Memin
- URL: http://fluid.irisa.fr

5.5. TYPHOON

GPU implementation of wavelet based motion estimator for Lidar data. This code is developped in coproperty between Inria and Chico.
FUNCTIONAL DESCRIPTION Typhoon is a motion estimation software specialized in fluid motion estimation. It is based on a dense optical flow technique associated to a multiscale wavelet representation of the estimated motion.

- Participants: Pierre Derian, Christopher Mauzey and Etienne Memin
- Partner: CSU Chico
- Contact: Etienne Memin
- URL: http://phys.csuchico.edu/lidar/typhoon/

5.6. GRT3D

FUNCTIONAL DESCRIPTION Reactive transport modeling has become an essential tool for understanding complex environmental problems. It is an important issue for MoMaS and C2S@EXA partners, in particular Andra. We have developed a method coupling transport and chemistry, based on a method of lines such that spatial discretization leads to a semi-discrete system of algebraic differential equations (DAE system). The main advantage is to use a complex DAE solver, which controls simultaneously the timestep and the convergence of Newton algorithm. The approach SIA uses a fixed-point method to solve the nonlinear system at each timestep, whereas the approach SNIA uses an explicit scheme.

- Participants: Yvan Crenner, Jocelyne Erhel
- Partner: ANDRA
- Contact: Jocelyne Erhel
- URL: http://www.irisa.fr/sage/recherche/chemistry.html

5.7. H2OLab

KEYWORDS: Simulation - Multiscale - Uncertainly - Heterogeneity - Hydrogeology - Groundwater - Contamination - Energy
SCIENTIFIC DESCRIPTION

The software platform contains a database which is interfaced through the web portal H2OWeb. It contains also software modules which can be used through the interface H2OGuilde. The platform H2OLab is an essential tool for the dissemination of scientific results. Currently, software and database are shared by the partners of the h2mno4 project.
FUNCTIONAL DESCRIPTION

The software platform H2OLab is devoted to stochastic simulations of groundwater flow and contaminant transport in highly heterogeneous porous and fractured geological media.

-Modeling and numerical simulation of aquifers -Porous and fractured heterogeneous media -Flow with mixed finite elements -Solute transport with a Lagrangian method -Stochastic modeling for data uncertainty.

- Participants: Jean-Raynald De Dreuzy, Jocelyne Erhel
- Partners: Université de Rennes 1 - CNRS - Université de Lyon - Université de Poitiers
- Contact: Jocelyne Erhel
- URL: http://h2olab.inria.fr/

5.8. PALMTREE

FUNCTIONAL DESCRIPTION

We present an easy-to-use package for the parallelization of Lagrangian methods for partial differential equations. In addition to the reduction of computation time, the code aims at satisfying three properties:

simplicity: the user just has to add the algorithm governing the behaviour of the particles. portability: the possibility to use the package with any compiler and OS. action-replay: the ability of the package to replay a selected batch of particles.

The last property allows the user to replay and capture the whole sample path for selected particles of a batch. This feature is very useful for debugging and catching some relevant information.

- Participants: Lionel Lenotre
- Contact: Jocelyne Erhel

6. New Results

6.1. Fluid motion estimation

6.1.1. Stochastic uncertainty models for motion estimation

Participants: Shengze Cai, Etienne Mémin, Musaab Khalid Osman Mohammed.

The objective consists here in relying on a stochastic transport formulation to propose a luminance conservation assumption dedicated to the measurement of large-scale fluid flows velocity. This formulation has the great advantage to incorporate from the beginning an uncertainty on the motion measurement. This uncertainty modeled as a possibly inhomogeneous random field uncorrelated in time can be estimated jointly to the motion estimates. Such a formulation, besides providing estimates of the velocity field and of its associated uncertainties, allows us to naturally define a linear multiresolution scale-space framework. It provides also a reinterpretation, in terms of uncertainty, of classical regularization functionals proposed in the context of motion estimation. This estimator, which extend a local motion estimator previously proposed in the team, has shown to improve significantly the results of the corresponding deterministic estimator. This kind method is assessed in the context of river hydrologics applications through a collaboration with an Irstea Lyon research group (HHLY). This study is performed within the PhD thesis of Musaab Mohammed.

6.1.2. Development of an image-based measurement method for large-scale characterization of indoor airflows

Participants: Dominique Heitz, Etienne Mémin, Romain Schuster.

The goal is to design a new image-based flow measurement method for large-scale industrial applications. From this point of view, providing in situ measurement technique requires the development of precise models relating the large-scale flow observations to the velocity, appropriate large-scale regularization strategies, and adapted seeding and lighting systems, like Hellium Filled Soap Bubles (HFSB) and led ramp lighting.This work conducted within the PhD of Romain Schuster in collaboration with the compagny ITGA has started in february 2016. The first step has been to evaluate the performances of a stochastic uncertainty motion estimator when using large scale scalar images, like those obtained when seeding a flow with smoke.

6.1.3. 3D flows reconstruction from image data

Participants: Dominique Heitz, Cédric Herzet.

Our work focuses on the design of new tools for the estimation of 3D turbulent flow motion in the experimental setup of Tomo-PIV. This task includes both the study of physically-sound models on the observations and the fluid motion, and the design of low-complexity and accurate estimation algorithms.

This year, we keep on our investigation on the problem of efficient volume reconstruction. Our work takes place within the context of some modern optimization techniques. First, 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). In 2016, this work has led to the publication of a contribution in the international journal IOP Measurement Science and Technology.

On top of this work, we also focussed on another crucial step of the volume reconstruction problem, namely the pruning of the model. The pruning task consists in identifying some positions in the volume of interest which cannot contains any particle. Removing this position from the problem can then potentially allow for a dramatic dimensionality reduction. This year, we provide a methodological answer to this problem through the prism of the so-called "screening" techniques which have been proposed in the community of machine learning. In 2016, this work led to the publication of one contribution in the proceedings of the international conference on acoustics, speech and signal processing (ICASSP'16).

6.1.4. 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 sparsity inducing 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.

This year, we contributed to this research axis by deriving conditions of success for the algorithms mentioned above when the amplitudes of the nonzero coefficients in the sparse vector obey some decay. In a TomoPIV context, this decay corresponds to the fact that not all the particles in the fluid diffuse the same quantity of light (notably because of illumination or radius variation). In particular, we show that the standard coherence-based guarantees for OMP/OLS can be relaxed by an amount which depends on the decay of the nonzero coefficients. In 2016, our work has led to the publication of one paper in the journal IEEE Transactions on Information Theory.

We also investigated a new methodology to take sparsity into account into variational assimilation problems. We focussed on the problem of estimating of scalar transported by an unknown velocity field, when only low-resolution observations of the scalar are supposed to be available. The goal is to reconstruct both a high-resolution version of the scalar and the velocity field, assuming that these quantities admit a sparse decomposition in some proper frames. The associated optimization problem typically involves millions of variables and thus requires dedicated optimization procedures to be tractable. In 2016, we proposed a new assimilation scheme combining state-of-the-art optimization techniques (forward-backward propagation, ADMM, Attouch's procedure) to address this problem. Our algorithm is provably convergent while exhibiting a complexity per iteration evolving linearly with the problem's dimensions. This contribution has led to a journal publication in SIAM Journal on Imaging Science.

6.2. Tracking, Data assimilation and model-data coupling

6.2.1. Stochastic fluid flow dynamics under uncertainty

Participants: Pierre Derian, 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 [10]. 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. 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). This year we have more deeply explored several uncertainty representations of classical geophysical models for ocean and atmosphere. This study have led to very promising stochastic representation for the Quasi Geostophic approximation (QG) with noises of different energy.

6.2.2. 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

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

Participants: Pranav Chandramouli, Dominique Heitz, Etienne Mémin.

In this axis of work we are exploring the use of optimal control techniques for the coupling of Large Eddies Simulation (LES) techniques and 2D image data. The objective is to reconstruct a 3D flow from a set of simultaneous time resolved 2D image sequences visualizing the flow on a set of 2D plans enlightened with laser sheets. This approach 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. The longer-term goal is 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 [7]. 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 has been 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. The control of subgrid parameterizations will be the main objective of the PhD of Pranav Chandramouli that is just starting.

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

Participant: Etienne Mémin.

This study is focused on the coupling of a large scale representation of the flow dynamics built from the location uncertainty principle with image data of finer resolution. The velocity field at large scales is described as a regular smooth component whereas the complement component is a highly oscillating random velocity field defined on the image grid but living at all the scales. Following this route we have assessed the performance of an ensemble variational assimilation technique with direct image data observation. Preliminary encouraging results have been obtained for simulation under uncertainty of 1D and 2D shallow water models.

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

Participants: Mamadou Diallo, Cédric Herzet, Etienne Mémin, Valentin Resseguier.

During the PhD thesis of Valentin Ressguier we proposed a new decomposition of the fluid velocity in terms of a large-scale continuous component with respect to time and a small-scale non continuous random component. Within this general framework, an uncertainty based representation of the Reynolds transport theorem and Navier-Stokes equations can be derived, based on physical conservation laws. This physically relevant stochastic model has been applied in the context of the POD-Galerkin method. The pertinence of this reduced order model has been successfully assessed on several wake flows. This study has been published in two conference papers and one journal article.

On the other hand, we investigated the problem of reduced-model construction from partial observations. In this line of search, our contribution was twofold. We first proposed a Bayesian framework for the construction of reduced-order models from image data. Our framework enables to account for any prior information on the system to reduce and takes the uncertainties on the parameters of the model into account. Interestingly, the proposed approach reduces to some well-known model-reduction techniques when the observations are not partial (i.e., the observation operator can be inverted). Second, we provided a theoretical analysis of our methodology in a simplified context (namely, the observations are supposed to be noiseless linear combinations of the state of the system). This result provides worst-case guarantees on the reconstruction performance which can be achieved by a reduced model built from the data. These contributions have led to the publications of one contribution in the proceedings of the international conference on acoustics, speech and signal processing (ICASSP'16). A journal version of these contributions has been submitted.

6.3. Analysis and modeling of turbulent flows

6.3.1. Singular and regular solutions to the Navier-Stokes equations (NSE) and relative turbulent models

Participant: Roger Lewandowski.

The common thread of this work is the problem set by J. Leray in 1934 : does a regular solution of the Navier-Stokes equations (NSE) with a smooth initial data develop a singularity in finite time, what is the precise structure of a global weak solution to the Navier-Stokes equations, and are we able to prove any uniqueness result of such a solution. This is a very hard problem for which there is for the moment no answer. Nevertheless, this question leads us to reconsider the theory of Leray for the study of the Navier-Stokes equations in the whole space with an additional eddy viscosity term that models the Reynolds stress in the context of large-scale flow modelling. It appears that Leray's theory cannot be generalized turnkey for this problem, so that things must be reconsidered from the beginning. This problem is approached by a regularization process using mollifiers, and particular attention must paid to the eddy viscosity term. For this regularized problem and when the eddy viscosity has enough regularity, we have able to prove the existence of a global unique solution that is of class C^{∞} in time and space and that satisfies the energy balance. Moreover, when the eddy viscosity is of compact support in space, uniformly in time, we recently shown that this solution converges to a turbulent solution to the corresponding Navier-Stokes equations when, the regularizing parameter goes to 0. These results are described in a paper that will be soon submitted to the journal Archive for Rational Mechanics and Analysis (ARMA).

In the same direction, we also finalized a paper in collaboration with L. Berselli (Univ. Pisa, Italy) about the well known Bardina's turbulent model. In this problem, we consider the Helmholtz filter usually used within the framework of Large Eddy Simulation. We carry out a similar analysis, by showing in particular that no singularity occurs for Bardina's model.

Another study in collaboration with B. Pinier, P. Chandramouli and E. Memin has been undertaken. This work takes place within the context of the PhD work of B. Pinier. We considered the standard turbulent models involving the Navier-Stokes equations with an eddy viscosity that depends on the Turbulent Kinetic Energy (TKE), coupled with an addition equation for the TKE. The problem holds in a 3D bounded domain, with the Manning law at the boundary for the velocity. We have modeled a flux condition at the boundary for the TKE. We prove that with these boundary conditions, the resulting problem has a distributional solution. Then a serie of numerical tests is performed in a parallelepiped with a non trivial bottom, showing the accuracy of the model in comparison with a direct numerical simulation of the Navier-Stokes equations.

6.3.2. Turbulence similarity theory for the modeling of Ocean Atmosphere interface

Participants: Roger Lewandowski, Etienne Mémin, Benoit Pinier.

The Ocean Atmosphere interface plays a major role in climate dynamics. This interaction takes place in a thin turbulent layer. To date no sastifying universal models for the coupling of atmospheric and oceanic models exists. In practice this coupling is realized through empirically derived interaction bulks. In this study, corresponding to the PhD thesis of Benoit Pinier, we aim at exploring similarity theory to identify universal mean profile of velocity and temperature within the mixture layer. The goal of this work consists in exhibiting eddy viscosity models within the primitive equations. We will also explore the links between those eddy viscocity models and the subgrid tensor derived from the uncertainty framework studied in the Fluminance group. In that prospect, we have started to study the impact of the introduction of a random modeling of the friction velocity on the classical wall law expression.

6.3.3. 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.

6.3.4. Numerical and experimental image and flow database

Participants: Pranav Chandramouli, Dominique Heitz.

The goal was to design a database for the evaluation of the different techniques developed in the Fluminance group. The first challenge was to enlarge a database mainly based on two-dimensional flows, with three-dimensional turbulent flows. Synthetic image sequences based on homogeneous isotropic turbulence and on circular cylinder wake have been provided. These images have been completed with time resolved Particle Image Velocimetry measurements in wake and mixing layers flows. This database provides different realistic conditions to analyse the performance of the methods: time steps between images, level of noise, Reynolds number, large-scale images. The second challenge was to carried out orthogonal dual plane time resolved stereoscopic PIV measurements in turbulent flows. The diagnostic employed two orthogonal and synchronized stereoscopic PIV measurements to provide the three velocity components in planes perpendicular and parallel to the streamwise flow direction. These temporally resolved planar slices observations will be used in 4DVar assimilation technique, integrating Direct Numerical Simulation (DNS) and Large Eddies Simulation (LES), to reconstruct three-dimensional turbulent flows. This reconstruction will be conducted within the PhD of Pranav Chandramouli. The third challenge was to carried out a time resolved tomoPIV experiments in a turbulent wake flow. 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.

6.4. Visual servoing approach for fluid flow control

6.4.1. Closed-loop control of a spatially developing shear layer

Participants: Christophe Collewet, Johan Carlier.

This study aims at controling one of the prototypical flow configurations encountered in fluid mechanics: the spatially developing turbulent shear layer occuring between two parallel incident streams with different velocities. Our goal is to maintain the shear-layer in a desired state and thus to reject upstream perturbations. As in all our previous works in flow control, we propose a vision-based approach to control this flow. We investigate the use of an optimal control based on a reduced linearized state space model of the Navier-Stokes equations. A steady desired state was first considered leading to a linear time-invariant system. The main problem consists to maintain the flow in his desired state in presence of unknown perturbation. Different strategies have been evaluated for different types of actuators and different cost functions. Even if our control law is based on a linearized approach, its efficiency has been validated on a realistic numerical Navier-Stokes 3D solver. This work has been submitted to the 20th World Congress of the International Federation of Automatic Control (IFAC).

6.5. Reactive transport

6.5.1. Reactive transport in porous media

Participant: Jocelyne Erhel.

In many environmental applications, transport of solutes is coupled with chemical reactions, either kinetic or at equilibrium. These reactions involve not only solutes, but also sorbed species and minerals. The mathematical model is a coupled set of nonlinear partial algebraic differential equations. A classical approach is to discretize first in space then in time. Since the problem is rather stiff, explicit time discretization suffers from a drastic CFL-like condition. On the other hand, implicit schemes allow large timesteps during some periods of simulation. Implicit Euler scheme is often used for monotonicity properties. The Jacobian is computed from the transport operator and the chemical operator. We have designed such a global approach and implemented it in our software GRT3D. We have done numerical experiments on the benchmark MoMaS.

Publications: 2 conferences and one journal article [15], [20], [21]

Grant: H2MNO4

6.5.2. Reactive transport in fractured-porous media

Participants: Yvan Crenner, Benjamin Delfino, Jean-Raynald de Dreuzy, Jocelyne Erhel.

Even in small numbers, fractures must be carefully considered for the geological disposal of radioactive waste. They critically enhance diffusivity, speed up solute transport, extend mixing fronts and, in turn, modify the physicochemical conditions of reactivity around possible storage sites. Numerous studies addressing various applications (e.g. radioactive waste storage, CO2 sequestration, geothermal storage, hydrothermal alteration) have shown that fractures cannot be simply integrated within an equivalent porous medium. Our objective is to develop a reactive transport model based on the separation of the fracture and matrix domains, with diffusion conditions differing between the fracture and in the matrix, appropriate flow-rock interactions at equilibrium in the matrix and fracture-matrix exchange conditions at their interface.

This year, we developed a numerical model for a chemical system with several minerals, which is representative of a storage site.

Publications: 2 conferences [28], [27]

Grant: ANDRA

6.6. Linear solvers

6.6.1. Sparse linear solvers

Participants: Jocelyne Erhel, David Imberti.

Sparse linear systems arise in computational science and engineering. The goal is to reduce the memory requirements and the computational cost, by means of high performance computing algorithms. We introduce a new variation on s-step GMRES in order to improve its stability, reduce the number of iterations necessary to ensure convergence, and thereby improve parallel performance. In doing so, we develop a new block variant that allows us to express the stability difficulties in s-step GMRES more fully.

Grants and projects: EXA2CT 8.2.1, EoCoE 8.2.2, C2S@EXA 8.1.7

Publications: 3 conférences [22], [23], [39]

7. Bilateral Contracts and Grants with Industry

7.1. Bilateral Contracts with Industry

7.1.1. Contract CERSAT/IFREMER

Participants: Etienne Mémin, Valentin Resseguier.

duration 36 months. This partnership between Inria and Ifremer funds the PhD of Valentin Resseguier, which aims at studying image based data assimilation strategies for oceanic models incorporating random uncertainty terms. The goal targeted will consist in deriving appropriate stochastic version of oceanic model and on top of them to devise estimation procedures from noisy data to calibrate the associated subgrid models.

7.1.2. Contract inter Carno IFREMER Inria

Participants: Etienne Mémin, Thibaut Tronchin.

duration 36 months. This contract aims at proposing image-based tools for the analysis of the hydraulic load of an immerged body. This project takes place within an inter Carnot cooperation between Ifremer and Inria.

7.1.3. Contract ITGA

Participants: Dominique Heitz, Etienne Mémin.

duration 36 months. This partnership between Inria, Irstea and ITGA funds the PhD of Romain Schuster. The goal of this CIFRE PhD is to design new image-based flow measurement methods for the study of industrial fluid flows. Those techniques will used in particular to calibrate industrial fume hood.

7.1.4. ANDRA project

Participants: Yvan Crenner, Benjamin Delfino, Jean-Raynald de Dreuzy, Jocelyne Erhel.

Contract with ANDRA (National Agency for Nuclear Waste)
Duration: three years from November 2015.
Title: reactive transport in fractured porous media
Coordination: Jocelyne Erhel.
Partners: Geosciences Rennes.
Abstract: Even in small numbers, fractures must be carefully considered for the geological disposal of radioactive waste. They critically enhance diffusivity, speed up solute transport, extend mixing fronts and, in turn, modify the physicochemical conditions of reactivity around possible storage sites. Numerous studies in various fields have shown that fractures cannot be simply integrated within an equivalent porous medium with a simple enhancement of its petro-physical properties (porosity and permeability). We propose a combined numerical and experimental approach to determine the influence on reactivity of typical fracture patterns found in some radioactive waste applications.

7.1.5. IFPEN project

Participants: Bastien Hamlat, Jocelyne Erhel.

Contract with IFPEN (Institut Français du Pétrole et Energies Nouvelles)
Duration: three years from October 2016.
Title: Fully implicit Formulations for the Simulation of Multiphase Flow and Reactive Transport
Coordination: Jocelyne Erhel.
Abstract: Modeling multiphase flow in porous media coupled with fluid-rock chemical reactions is essential in order to understand the origin of sub-surface natural resources and optimize their extraction. This project aims to determine optimal strategies to solve the coupled transport and chemical reaction equations describing the physical processes at work in reactive multiphase flow in porous media. Three different formulations show great potential to accurately solve these equations. Two are fully implicit ("Reactive Coats" and "Semi-smooth Newton)" and one is an operator splitting approach. These formulations are still incomplete at the moment. The work will focus on extending the existing formulations to more complex physical phenomena, study their stability, convergence and theoretical equivalence. Another objective is to provide practical solutions to efficiently solve the resulting non-linear systems.

8. Partnerships and Cooperations

8.1. National Initiatives

8.1.1. Comins'lab: SEACS : Stochastic modEl-dAta-Coupled representationS for the analysis, simulation and reconstruction of upper ocean dynamics

Participant: Etienne Mémin.

duration 48 months. The SEACS project whose acronym stands for: "Stochastic modEl-dAta-Coupled representationS for the analysis, simulation and reconstruction of upper ocean dynamics" is a Joint Research Initiative between the three Britanny clusters of excellence of the "Laboratoires d'Excellence" program: Cominlabs, Lebesgue and LabexMer centered on numerical sciences, mathemathics and oceanography respectively. Within this project we aim at studying the potential of large-scale oceanic dynamics modeling under uncertainty for ensemble forecasting and satellite image data assimilation.

8.1.2. ANR JCJC GERONIMO : Advanced GEophysical Reduced-Order Model construction from IMage Observations

Participant: Cédric Herzet.

duration 48 months. The GERONIMO project which started in March 2014 aims at devising new efficient and effective techniques for the design of geophysical reduced-order models from image data. The project both arises from the crucial need of accurate low-order descriptions of highly-complex geophysical phenomena and the recent numerical revolution which has supplied the geophysical scientists with an unprecedented volume of image data. The project is placed at the intersection of several fields of expertise (Bayesian inference, matrix factorization, sparse representations, etc.) which will be combined to handle the uncertainties associated to image measurements and to characterize the accurate reduced dynamical systems.

8.1.3. ANR BECOSE : Beyond Compressive Sensing: Sparse approximation algorithms for ill-conditioned inverse problems.

Participants: Dominique Heitz, Cédric Herzet.

duration 48 months. The BECOSE project aims to extend the scope of sparsity techniques much beyond the academic setting of random and well-conditioned dictionaries. In particular, one goal of the project is to step back from the popular L1-convexification of the sparse representation problem and consider more involved nonconvex formulations, both from a methodological and theoretical point of view. The algorithms will be assessed in the context of tomographic Particle Image Velocimetry (PIV), a rapidly growing imaging technique in fluid mechanics that will have strong impact in several industrial sectors including environment, automotive and aeronautical industries. The consortium gathers the Fluminance and Panama Inria research teams, the Research Center for Automatic Control of Nancy (CRAN), The Research Institute of Communication and Cybernetics of Nantes (IRCCyN), and ONERA, the French Aerospace Lab.

8.1.4. ANR-MN: H2MNO4 project

Participants: Yvan Crenner, Benjamin Delfino, Jean-Raynald de Dreuzy, Jocelyne Erhel, Lionel Lenôtre.

Contract with ANR, program Modèles Numériques
Duration: four years from November 2012.
Title: Original Optimized Object Oriented Numerical Model for Heterogeneous Hydrogeology.
Coordination: Jocelyne Erhel and Géraldine Pichot, with Fabienne Cuyollaa.
Partners: Geosciences Rennes, University of Poitiers, University of Lyon 1, Andra, Itasca.
International collaborations: University of San Diego (USA), UPC, Barcelona (Spain)
Web page: http://h2mno4.inria.fr/

Abstract: The project H2MNO4 develops numerical models for reactive transport in heterogeneous media. It defines six mathematical and computational challenges and three applications for environmental problems with societal impact. We presented a poster at the ANR-day (rencontre du numérique, Parius, Nov. 2016)

8.1.5. INSU-LEFE: Toward new methods for the estimation of sub-meso scale oceanic streams

Participant: Cédric Herzet.

duration 36 months. This project tackles the problem of deriving a precise submesoscale characterization of ocean currents from satellite data. The targeted methodologies should in particular enable the exploitation of data of different nature (for example sea surface temperature or height) and/or resolutions. This 36-month project benefits from a collaboration with the Laboratoire de Météorologie Dynamique, Ecole Normale Supérieure, Paris.

8.1.6. INSU-LEFE: MODELER

Participant: Etienne Mémin.

duration 24 months. This project with MeteoFrance aims at exploring error modeling and stochastic parameterization in geophysical flow dynamics. The theory explored in this context should enable the construction of unified image data assimilation strategies.

8.1.7. Inria Project Lab: C2S@EXA project

Participants: Yvan Crenner, Jocelyne Erhel.

Title: C2S@EXA - Computer and Computational Scienecs at Exascale
Duration: from January 2012 until April 2017
Coordination: S. Lanteri, Nachos team.
Partners: Inria teams working on HPC; external partners: ANDRA and CEA.
Webpage: http://www-sop.inria.fr/c2s_at_exa/
Abstract: The C2S@Exa Inria Project Lab is concerned with the development of numerical modeling methodologies that fully exploit the processing capabilities of modern massively parallel architectures in the context of a number of selected applications related to important scientific and technological challenges for the quality and the security of life in our society. The team participated in several workshops.

8.1.8. GENCI: project on advanced linear solvers

Participants: Yvan Crenner, Jocelyne Erhel, David Imberti.

Title: Numerical models for hydrogeology
Duration: 2016
Coordination: J. Erhel
Webpage: http://www.genci.fr/
Abstract: To run large scale simulations, we defined a project, based on the platform H2OLab and on a new GMRES solver. We obtained and used computing time on machines located at GENCI supercomputing centers.

8.1.9. GDR MANU

Participants: Benjamin Delfino, Jocelyne Erhel.

Title: Mathematics for Nuclear industry
Duration: From 2016 to 2019
Coordination: C. Cancès
Webpage: http://gdr-manu.math.cnrs.fr/
Abstract: The working group MANU is a follow-up to the group MOMAS. It covers many subjects related to mathematical modeling and numerical simulations for problems arising from nuclear industry and nuclear waste disposal. The team participated in the conference JEMP2016.

8.2. European Initiatives

8.2.1. EXA2CT

Participants: Jocelyne Erhel, David Imberti.

Title: EXascale Algorithms and Advanced Computational Techniques

Programm: FP7

Duration: September 2013 - August 2016

Coordinator: S. Ashby, IMEC, Belgium

Partners:

Fraunhofer-Gesellschaft Zur Foerderung Der Angewandten Forschung E.V (Germany)

Interuniversitair Micro-Electronica Centrum Vzw (Belgium)

Intel Corporations (France)

Numerical Algorithms Group Ltd (United Kingdom)

Systems Solutions for Research (Germany)

Universiteit Antwerpen (Belgium)

Universita della Svizzera italiana (Switzerland)

Universite de Versailles Saint-Quentin-En-Yvelines. (France)

Vysoka Skola Banska - Technicka Univerzita Ostrava (Czech Republic)

Inria contact: Luc Giraud

Abstract: Numerical simulation is a crucial part of science and industry in Europe. The advancement of simulation as a discipline relies on increasingly compute intensive models that require more computational resources to run. This is the driver for the evolution to exascale. The EXA2CT project brings together experts at the cutting edge of the development of solvers, related algorithmic techniques, and HPC software architects for programming models and communication.

8.2.2. EOCOE project

Participant: Jocelyne Erhel.

Program: EINFRA-5-2015

Project acronym: EoCoE

Project title: Energy oriented Center of Excellence for computer applications

Duration: 36 months

Coordinator: CEA

Other partners: organisme, labo (pays) : 12 other partners

Abstract: the EoCoE objectives aims at firstly, to design, test and spread new methodological and organisational paradigms (Objectives 1, 3, and 4) driven by the users communities and, secondly, to contribute to mathematical and computer sciences challenges on the whole HPC tool chain (Objective 2).

8.3. International Initiatives

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

8.3.1.1. LFD-FLU

Title: Large-scale Fluid Dynamics analysis from FLow Uncertainty

International Partner (Institution - Laboratory - Researcher):

Universidad de Buenos Aires (Argentina) - Department of Computer Science and Electrical Engineering - Guillermo Artana

Start year: 2016

See also: http://www.irisa.fr/prive/memin/LFD-FLU/

The first objective of this associate team is primarily concerned with the establishment of efficient fluid flow image data analysis procedures. This concerns for instance data assimilation issues to reconstruct meaningful numerical representation of experimental fluid flows for analysis purpose. The second objective focuses on the incorporation of uncertainties in the flow dynamical evolution models

8.3.2. Inria International Partners

8.3.2.1. Informal International Partners

Imperial College (London, UK) We have initiated a collaboration with the Department of Aeronautics within the PhD thesis of Pranav Chandramoulli

Chico California State University (USA), We have pursue our collaboration with the group of Shane Mayor on the GPU implementation of wavelet based motion estimator for Lidar data. This code is developped in coproperty between Inria and Chico.

College of Control Science & Engineering of Zhejiang University We have initiated a collaboration with Prof. Chao Xu on the study of fluid motion estimator.

8.4. International Research Visitors

8.4.1. Visits of International Scientists

- 2 month sojourn of Gisela Charo (PhD student University of Buenos Aires) to work with Etienne Mémin and Valentin Resseguier within the associate team LFD
- 2 weeks visit of Alejandro Gronskis (Researcher Conicet Argentina) to work with Dominique Heitz, Etienne Mémin and Pranav Chandramouli within the associate team LFD
- Sojourn of 9 month of Shengze Cai PhD student in the College of Control Science & Engineering, Zhejiang University to work with Etienne Mémin
- 2 weeks visit of Prof. Luigi Berselli (U. Pisa) to work with Roger Lewandowski.

9. Dissemination

9.1. Promoting Scientific Activities

9.1.1. Scientific Events Organisation

9.1.1.1. General Chair, Scientific Chair

- J. Erhel organizes with J-R. de Dreuzy and T. Le Borgne the international conference CWMR (Saint-Malo, France, June 2018).

9.1.1.2. Member of the Organizing Committees

Etienne Mémin

- SWGEN (Vannes) program committee
- Scientific committee of the national colloqium on data assimilation, (Grenoble)

9.1.2. Scientific Events Selection

9.1.2.1. Member of the Conference Program Committees

Jocelyne Erhel

- international advisory committee of the parallel CFD conferences (Kobe, Japan, May 2016).
- program committee of the international conference CARI 2016.
- scientific committee of JEMP 2016.
- scientific committee of NLAA 2016.
- program committee of the workshop Visualization in Environmental Sciences 2016 (co-event of EuroVis)

9.1.3. Journal

9.1.3.1. Member of the Editorial Boards

Jocelyne Erhel

- member of the editorial board of ETNA.
- member of the editorial board of ESAIM:Proceedings and Surveys.

Etienne Mémin

- Associate editor for the Int. Journal of Computer Vision (IJCV)
- Associate editor for the Image and Vision Computing Journal (IVC)

9.1.3.2. Reviewer - Reviewing Activities

Jocelyne Erhel: Reviewer for the journals ADWR, ARIMA, JCAM, MATCOM
Dominique Heitz: Reviewer for Exp. in Fluids, ASME J. on Heat Transfer
Cédric Herzet: Reviewer for IEEE Tr. on Signal Processing, IEEE Tr. on Information Theory
Etienne Mémin: Reviewer for Tellus-A, IEEE Im. Proc., IEEE trans. Pat. Anal. Mach. Intel. , Im. Vis. Comp., Exp. in Fluids, Nonlinear Proc. in Geophysics., Journ. of Comp. Phys, Fluid Dynamics Research.

9.1.4. Invited Talks

Dominique Heitz

- Next generation transport aircraft workshop, Honolulu, Hawai, 22-25, February, 2016.
- Cap Aliment training "Les technologies douces de conservation des denrées alimentaires", Nantes, 13, October, 2016. Assisses du Génie des Procédés, région OUEST - Nantes, 9, november, 2016

Cédric Herzet

- GdR Isis « Algorithmes gloutons pour l'optimisation sous contrainte de parcimonie », Juin 2016

Roger Lewandowski

- Special Session on Above and Beyond Fluid Flow studies: In celebration of the 60th birthday of Prof. William Layton » within the Fall Western Sectional Meeting of the AMS, University of Denver, Denver, CO October 8-9, 2016.

Etienne Mémin

- E. Mémin. Représentation sous incertitude d'écoulements géophysiques, Huitième Ecole Interdisciplinaire de Rennes sur les Systèmes Complexes, Oct. 2016.

9.1.5. Leadership within the Scientific Community

- J. Erhel is scientific coordinator of the website Interstices (since June 2012). https://interstices.info.

9.1.6. Scientific Expertise

- J. Erhel is a member of the scientific council of IFPEN, since April 2016.
- J. Erhel was reviewer for ANR.

9.1.7. Research Administration

Jocelyne Erhel

- correspondent of Maison de la Simulation for Inria Rennes.
- correspondent of AMIES for Inria Rennes, from September 2015.
- member of the Inria national committee for secondment, 2016.
- member of the Inria local committee for health and safety (référent chercheur) from January 2016.
- member of the Inria administrative commission (CAP) for researchers, from January 2016.

Dominique Heitz

- Responsible of the Irstea ACTA Team
- Member of Pôle Cristal scientific council
- Member of Irstea OPAALE research unit Executive Committee
- Member of Irstea center of Rennes Executive Committee

Etienne Mémin

- Responsible of the "Commission Développement Technologique" Inria Rennes
- Member of the "Commission Personnel" Inria-IRISA Rennes

9.2. Teaching - Supervision - Juries

9.2.1. Teaching

Licence: Jocelyne Erhel, Optimisation, 24h, niveau L3, ENSAI Rennes

Licence : Dominique Heitz, Mécanique des fluides, 30h, niveau L2 INSA Rennes

Master: Jocelyne Erhel, modélisation et calcul scientifique, 12h, niveau M2, INSA Rennes

Master: Jocelyne Erhel, arithmétique flottante, 4h, niveau M1, INSA Rennes

Master : Dominique Heitz, Mécanique des fluides, 25h, niveau M1, Dep GMA INSA Rennes

Master : Cédric Herzet, Représentations parcimonieuses et compressed sensing, niveau M2, ENSAI, 12h

Master : Cédric Herzet, Représentations parcimonieuses et compressed sensing, niveau M2, niveau M2, INSA, 10h

Master: Roger Lewandowski, Euler and the Navier-Stokes equations, M2, master « fondamental mathematics ».

Master : Etienne Mémin, Analyse du mouvement, Mastere Informatique, 15h, niveau M2, Université de Rennes 1.

Master : Etienne Mémin, Vision par ordinateur , 15h, niveau M2, ESIR Université de Rennes 1.

9.2.2. Supervision

Master research work: C. Bonvoisin (ENS Cachan) February 2016 to June 2016, advisor R. Lewandowski

PhD in progress: B. Delfino, University of Rennes 1, November 2015, co-advisors J.-R. de Dreuzy and J. Erhel.

PhD in progress: P.-M. Gibert, University of Lyon, October 2015, co-advisors D. Tromeur-Dervout and J. Erhel.

PhD in progress: B. Hamlat, University of Rennes 1, October 2016, co-advisors J. Erhel and A. Michel.

PhD in progress : Benoit Pinier, Scale similarity and uncertainty for Ocean-Atmosphere coupled models, started 01/10/2014, supervisors: Roger Lewandowski, Etienne Mémin

PhD in progress : Valentin Resseguier, Oceanic models under uncertainty and image assimilitation, started 01/10/2013, Bertrand Chapron (Ifremer), Etienne Mémin

PhD in progress : Pranav Chandamouli, Turbulent complex flows reconstruction via data assimilation in large eddy models, started october 2015, Dominique Heitz, Etienne Mémin.

PhD in progress : Romain Schuster, Large-scale fluid motion estimation, started october 2016, Dominique Heitz, Etienne Mémin.

9.2.3. Juries

Jocelyne Erhel

- Rafife Nheili, PhD, Univ. Perpignan (rapporteur)

Etienne Mémin

- Laurent Cordier HDR, Univ. Poitiers.
- Nicolas Papadakis, HDR IMB, Univ. Bordeaux
- Yann Michel, HDR Meteofrance, Univ. Toulouse Paul Sabatier (Rapporteur)
- Van Linh Nguyen, PhD Univ. Lille (Rapporteur)
- Iliass Azijli, PhD TU Delft (Rapporteur),
- Raphael Legrand, PhD Meteofrance Univ. Paul Sabatier Toulouse (Rapporteur)

9.3. Popularization

Jocelyne Erhel

- présidente du jury du rallye de mathématiques du CNED, 2016.
- talk at day "au coeur des maths, enfermement ou liberté ?", LVN, May 2016.
- scientific responsible of the scientific culture web journal "Interstice" (https://interstices.info/)

Dominique Heitz

- Interview dans L'Usine Nouvelle, No 3479, pp. 30-31, 2016

10. Bibliography

Major publications by the team in recent years

[1] K. BURRAGE, J. ERHEL.*On the performance of various adaptive preconditioned GMRES*, in "Numerical Linear Algebra with Applications", 1998, vol. 5, p. 101-121.

[2] J. CARRAYROU, J. HOFFMANN, P. KNABNER, S. KRÄUTLE, C. DE DIEULEVEULT, J. ERHEL, J. VAN DER LEE, V. LAGNEAU, K. MAYER, K. MACQUARRIE.*Comparison of numerical methods for simulating strongly non-linear and heterogeneous reactive transport problems. The MoMaS benchmark case*, in "Computational Geosciences", 2010, vol. 14, n° 3, p. 483-502.

[3] R. CHOQUET, J. ERHEL.*Newton-GMRES algorithm applied to compressible flows*, in "International Journal for Numerical Methods in Fluids", 1996, vol. 23, p. 177-190.

[4] T. CORPETTI, D. HEITZ, G. ARROYO, E. MÉMIN, A. SANTA-CRUZ.*Fluid experimental flow estimation based on an optical-flow scheme*, in "Experiments in fluids", 2006, vol. 40, p. 80–97.

[5] J.-R. DE DREUZY, A. BEAUDOIN, J. ERHEL.*Asymptotic dispersion in 2D heterogeneous porous media determined by parallel numerical simulations*, in "Water Resource Research", 2007, vol. 43, n° W10439, doi:10.1029/2006WR005394.

[6] J. ERHEL, S. RAULT.*Algorithme parallèle pour le calcul d'orbites : Parallélisation à travers le temps*, in "Technique et science informatiques", 2000, vol. 19, n° 5.

[7] A. GRONSKIS, D. HEITZ, E. MÉMIN.*Inflow and initial conditions for direct numerical simulation based on adjoint data assimilation*, in "Journal of Computational Physics", 2013, vol. 242, p. 480-497 [*DOI :* 10.1016/J.JCP.2013.01.051], http://www.sciencedirect.com/science/article/pii/S0021999113001290.

[8] D. HEITZ, E. MÉMIN, C. SCHNOERR.*Variational Fluid Flow Measurements from Image Sequences: Synopsis and Perspectives*, in "Experiments in fluids", 2010, vol. 48, n° 3, p. 369–393.

[9] H. HOTEIT, J. ERHEL, R. MOSÉ, B. PHILIPPE, P. ACKERER.*Numerical Reliability for Mixed Methods Applied to Flow Problems in Porous Media*, in "Computational Geosciences", 2002, vol. 6, p. 161-194.

[10] E. MÉMIN.*Fluid flow dynamics under location uncertainty*, in "Geophysical & Astrophysical Fluid Dynamics", 2014, vol. 108, n° 2, p. 119-146, http://dx.doi.org/10.1080/03091929.2013.836190.

[11] N. PAPADAKIS, E. MÉMIN.*A variational technique for time consistent tracking of curves and motion*, in "Journal of Mathematical Imaging and Vision", 2008, vol. 31, n° 1, p. 81–103, http://www.irisa.fr/fluminance/publi/papers/Papadakis-Memin-JMIV07.pdf.

[12] Y. SAAD, M. YEUNG, J. ERHEL, F. GUYOMARC'H.*A deflated version of the Conjugate Gradient Algorithm*, in "SIAM Journal on Scientific Computing", 2000, vol. 21, n° 5, p. 1909-1926.

Publications of the year

Articles in International Peer-Reviewed Journal

[13] P. ARBOGAST, O. PANNEKOUCKE, L. RAYNAUD, R. LALANNE, E. MÉMIN.*Object-oriented processing of CRM precipitation forecasts by stochastic filtering*, in "Quarterly Journal of the Royal Meteorological Society", 2016 [*DOI :* 10.1002/QJ.2871], https://hal.inria.fr/hal-01378366.

[14] I. BARBU, C. HERZET.*A new approach for volume reconstruction in TomoPIV with the alternating direction method of multipliers*, in "Measurement Science and Technology", 2016, vol. 27, 21 [*DOI :* 10.1088/0957-0233/27/10/104002], https://hal.inria.fr/hal-01419329.

[15] J. ERHEL, S. SABIT.*Analysis of a global reactive transport model and results for the MoMaS benchmark*, in "Mathematics and Computers in Simulation", 2017, https://hal.inria.fr/hal-01405698.

[16] C. HERZET, A. DRÉMEAU, C. SOUSSEN.*Relaxed recovery conditions for OMP/OLS by exploiting both coherence and decay*, in "IEEE Transactions on Information Theory", January 2016, vol. 62, n° 1, p. 459-470 [*DOI :* 10.1109/TIT.2015.2490660], https://hal.inria.fr/hal-01058077.

[17] C. HERZET, A. DRÉMEAU, C. SOUSSEN.*Relaxed Recovery Conditions for OMP/OLS by Exploiting Both Coherence and Decay*, in "IEEE Transactions on Information Theory", 2016, vol. 62, p. 459 - 470 [*DOI :* 10.1109/TIT.2015.2490660], https://hal.inria.fr/hal-01419333.

[18] P. HÉAS, A. DRÉMEAU, C. HERZET.*An Efficient Algorithm for Video Super-Resolution Based On a Sequential Model*, in "SIAM Journal on Imaging Sciences", 2016, vol. 9, n° 2, p. 537-572, 37 pages [*DOI :* 10.1137/15M1023956], https://hal.inria.fr/hal-01158551.

[19] S. NAKHOSTIN, H. CLENET, T. CORPETTI, N. COURTY.*Joint Anomaly Detection and Spectral Unmixing for Planetary Hyperspectral Images*, in "IEEE Transactions on Geoscience and Remote Sensing", 2016, vol. 54, p. 6879 - 6894 [*DOI :* 10.1109/TGRS.2016.2586188], https://hal.archives-ouvertes.fr/hal-01377230.

Invited Conferences

[20] J. ERHEL.*Un modèle mathématique de réactions chimiques et de transport réactif*, in "Calcul Scientifique et Modélisation Mathématique", Amiens, France, 2016, https://hal.inria.fr/hal-01405739.

[21] J. ERHEL, S. SABIT, T. MIGOT.*Reactive transport simulations using a global approach*, in "computational methods in water resources (CMWR)", Toronto, Canada, 2016, https://hal.inria.fr/hal-01405708.

[22] D. IMBERTI, J. ERHEL.*Parallel GMRES using a variable s-step Krylov basis*, in "SIAM Conference on Parallel Processing for scientific computing", Paris, France, 2016, https://hal.inria.fr/hal-01405733.

[23] D. IMBERTI, J. ERHEL.*Varying the s in s-step GMRES*, in "Numerical Linear Algebra and Applications (NL2A)", Marseille, France, 2016, https://hal.inria.fr/hal-01405735.

International Conferences with Proceedings

[24] C. HERZET, A. MALTI.*Safe screening tests for lasso based on firmly non-expansiveness*, in "IEEE International conference on acoustics, speech and signal processing (ICASSP'16)", Shanghai, China, March 2016, https://hal.inria.fr/hal-01419314.

[25] P. HÉAS, C. HERZET.*Reduced-Order Modeling of Hidden Dynamics*, in "IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASPP)", Shangai, China, March 2016, p. 1268–1272 [*DOI :* 10.1109/ICASSP.2016.7471880], https://hal.inria.fr/hal-01246074.

Conferences without Proceedings

[26] N. COURTY, R. FLAMARY, D. TUIA, T. CORPETTI.*Optimal Transport for Data Fusion in Remote Sensing*, in "IGARSS", Beijing, China, July 2016, https://hal.archives-ouvertes.fr/hal-01377234.

[27] B. DELFINO, J.-R. DE DREUZY, J. ERHEL, B. COCHEPIN, Y. MÉHEUST.*Impact of fractures on diffusively dominated reactive transport: application to radioactive waste storage studies*, in "Computational Methods for Water Resources (CMWR)", Toronto, Canada, 2016, https://hal.inria.fr/hal-01405716.

[28] B. DELFINO, J. R. DE DREUZY, J. ERHEL.*Impact of fractures on diffusion dominated reactive transport in porous media: application to the study of a radioactive waste storage*, in "13èmes Journéess d'études des Milieux Poreux 2016", Anglet, France, October 2016, https://hal.archives-ouvertes.fr/hal-01394582.

[29] S. NAKHOSTIN, N. COURTY, R. FLAMARY, T. CORPETTI.*Supervised Planetary Unmixing with Optimal Transport*, in "WHISPERS 8th workshop on Hyperspectral image and Signal Processing: Evolution in Remote Sensing", Los Angeles, United States, August 2016, https://hal.archives-ouvertes.fr/hal-01377236.

[30] V. RESSEGUIER, E. MÉMIN, B. CHAPRON, P. DÉRIAN.*Stochastic parameterization of geophysical flows through modelling under location uncertainty*, in "Data Analysis and Modeling in Earth Sciences (DAMES) 2016", Hambourg, Germany, September 2016, https://hal.inria.fr/hal-01377719.

[31] V. RESSEGUIER, E. MÉMIN, B. CHAPRON.*Likely chaotic transitions of large-scale fluid flows using a stochastic transport model*, in "9th Chaotic Modeling and Simulation International Conference (CHAOS2016)", Londres, United Kingdom, May 2016, https://hal.inria.fr/hal-01377702.

[32] V. RESSEGUIER, E. MÉMIN, B. CHAPRON.*Randomized fluid dynamics based on subgrid transport*, in "Workshop on Stochastic Weather Generators", Vannes, France, May 2016, https://hal.inria.fr/hal-01377747.

[33] V. RESSEGUIER, E. MÉMIN, B. CHAPRON.*Transport along stochatic flows in fluid dynamics*, in "Workshop - Statistical methods for dynamical stochastic models - DYNSTOCH 2016", Rennes, France, June 2016, https://hal.inria.fr/hal-01377723.

Other Publications

[34] I. BARBU, A. GUIBERT, P. LOISEL, D. HEITZ.*Optical Flow Estimation out of Infrared Image Sequences of Biogas*, June 2016, The 17th International Symposium on Flow Visualization, Poster [*DOI :* 10.13140/RG.2.2.35868.28808], https://hal.archives-ouvertes.fr/hal-01416030.

[35] I. BARBU, A. GUIBERT, P. LOISEL, D. HEITZ.*Validation d'une technique de vélocimétrie du gaz par imagerie infrarouge*, September 2016, CFTL2016, Poster [*DOI :* 10.13140/RG.2.2.11095.68006], https://hal.archives-ouvertes.fr/hal-01417131.

[36] A. DRÉMEAU, C. HERZET.*DOA estimation in structured phase-noisy environments: technical report*, 2016, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01394048.

[37] C. HERZET, A. DRÉMEAU, P. HÉAS.*Model Reduction from Partial Observations*, 2016, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01394059.

[38] P. HÉAS, C. HERZET.*Low-rank Approximation and Dynamic Mode Decomposition*, 2016, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01394064.

[39] D. IMBERTI, J. ERHEL.*Vary the s in Your s-step GMRES*, July 2016, working paper or preprint, https://hal.inria.fr/hal-01299652.

[40] S. KADRI HAROUNA, E. MÉMIN.*Stochastic representation of the Reynolds transport theorem: revisiting large-scale modeling*, November 2016, working paper or preprint, https://hal.inria.fr/hal-01394780.

[41] V. RESSEGUIER, E. MÉMIN, B. CHAPRON.*Geophysical flows under location uncertainty, Part I Random transport and general models*, October 2016, working paper or preprint, https://hal.inria.fr/hal-01391420.

[42] V. RESSEGUIER, E. MÉMIN, B. CHAPRON.*Geophysical flows under location uncertainty, Part II Quasi-geostrophy and efficient ensemble spreading*, October 2016, working paper or preprint, https://hal.inria.fr/hal-01391476.

[43] V. RESSEGUIER, E. MÉMIN, B. CHAPRON.*Geophysical flows under location uncertainty, Part III SQG and frontal dynamics under strong turbulence conditions*, October 2016, working paper or preprint, https://hal.inria.fr/hal-01391484.

[44] V. RESSEGUIER, E. MÉMIN, B. CHAPRON.*Oceanic models under uncertainty*, June 2016, CMG 2016 : 31st IUGG Conference on Mathematical Geophysics, Poster, https://hal.inria.fr/hal-01377710.

[45] V. RESSEGUIER, E. MÉMIN, B. CHAPRON.*Stochastic subgrid tensor for geophysical flow modeling*, July 2016, 8th European Postgraduate Fluid Dynamics Conference, Poster, https://hal.inria.fr/hal-01377741.

[46] V. RESSEGUIER, E. MÉMIN, D. HEITZ, B. CHAPRON.*Stochastic modelling and diffusion modes for POD models and small-scale flow analysis*, November 2016, working paper or preprint, https://hal.inria.fr/hal-01400119.

References in notes

[47] R. ADRIAN.*Particle imaging techniques for experimental fluid mechanics*, in "Annal Rev. Fluid Mech.", 1991, vol. 23, p. 261-304.

[48] T. BEWLEY.*Flow control: new challenges for a new Renaissance*, in "Progress in Aerospace Sciences", 2001, vol. 37, p. 21–58.

[49] S. BEYOU, A. CUZOL, S. GORTHI, E. MÉMIN.*Weighted Ensemble Transform Kalman Filter for Image Assimilation*, in "TellusA", January 2013, vol. 65, n° 18803.

[50] H. CHOI, P. MOIN, J. KIM.*Direct numerical simulation of turbulent flow over riblets*, in "Journal of Fluid Mechanics", 1993, vol. 255, p. 503–539.

[51] T. CORPETTI, E. MÉMIN.*Stochastic Uncertainty Models for the Luminance Consistency Assumption*, in "IEEE Trans. Image Processing", 2012, vol. 21, n° 2, p. 481–493.

[52] G. EVENSEN.*Sequential data assimilation with a non linear quasi-geostrophic model using Monte Carlo methods to forecast error statistics*, in "J. Geophys. Res.", 1994, vol. 99 (C5), n° 10, p. 143–162.

[53] J. FAVIER.*Contrôle d'écoulements : approche expérimentale et modélisation de dimension réduite*, Institut National Polytechnique de Toulouse, 2007.

[54] J. FAVIER, A. KOURTA, G. LEPLAT.*Control of flow separation on a wing profile using PIV measurements and POD analysis*, in "IUTAM Symposium on Flow Control and MEMS", London, UK, September 19-22, 2006.

[55] N. GORDON, D. SALMOND, A. SMITH.*Novel approach to non-linear/non-Gaussian Bayesian state estimation*, in "IEEE Processing-F", April 1993, vol. 140, n^o 2.

[56] A. GUÉGAN, P. SCHMID, P. HUERRE.*Optimal energy growth and optimal control in swept Hiemenz flow*, in "J. Fluid Mech.", 2006, vol. 566, p. 11–45.

[57] B. HORN, B. SCHUNCK.*Determining Optical Flow*, in "Artificial Intelligence", August 1981, vol. 17, n^o 1-3, p. 185–203.

[58] F.-X. LE DIMET, O. TALAGRAND.*Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects*, in "Tellus", 1986, n^o 38A, p. 97–110.

[59] J. LIONS.*Optimal Control of Systems Governed by Partial Differential Equations*, Springer-Verlag, 1971.

[60] L. MATHELIN, O. LE MAÎTRE.*Robust control of uncertain cylinder wake flows based on robust reduced order models*, in "Computer and Fluids", 2009, vol. 38, p. 1168–1182.

[61] B. PROTAS, J. WESFREID.*Drag force in the open-loop control of the cylinder wake in the laminar regime*, in "Physics of Fluids", February 2002, vol. 14, n^o 2, p. 810–826.

[62] I. WYGNANSKI.*Boundary layer flow control by periodic addition of momentum*, in "4th AIAA Shear Flow Control Conference", USA, June 29-July 2, 1997.

Project-Team GENSCALE

Scalable, Optimized and Parallel Algorithms for Genomics

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

IN PARTNERSHIP WITH:
CNRS
Ecole normale supérieure de Cachan
Université Rennes 1

RESEARCH CENTER
Rennes - Bretagne-Atlantique

THEME
Computational Biology

Table of contents

Project-Team GENSCALE

Creation of the Team: 2012 January 01, updated into Project-Team: 2013 January 01

Keywords:

Computer Science and Digital Science:

3.1.2. - Data management, quering and storage
3.1.8. - Big data (production, storage, transfer)
3.3.2. - Data mining
3.3.3. - Big data analysis
7.1. - Parallel and distributed algorithms
7.3. - Optimization

Other Research Topics and Application Domains:

1.1.6. - Genomics
1.1.9. - Bioinformatics
2.2.3. - Cancer

1. Members

Research Scientists

Dominique Lavenier [Team leader, CNRS, Senior Researcher, HDR]
Claire Lemaitre [Inria, Researcher]
Pierre Peterlongo [Inria, Researcher, HDR]
Guillaume Rizk [Associate Researcher, ANR HydroGen]

Faculty Members

Roumen Andonov [Univ. Rennes I, Professor, HDR]
Antonio Mucherino [Univ. Rennes I, Associate Professor, until Aug 2016]

Technical Staff

Fabrice Legeai [Inra]
Laurent Bouri [CNRS]
Jennifer Del Giudice [Inria, from Mar 2016]
Patrick Durand [Inria]
Sebastien Letort [CNRS]
Ivaylo Petrov [Inria, until Jan 2016]
Chloe Riou [Inria, until Apr 2016]
Charles Deltel [Inria, Research engineer, 50% time dedicated to the GenScale project]

PhD Students

Gaetan Benoit [Inria, granted by Hydrogen ANR project]
Sebastien Francois [Univ. Rennes I]
Cervin Guyomar [Univ. Rennes I]
Antoine Limasset [Univ. Rennes I]
Camille Marchet [Univ. Rennes I]
Hoang Son Pham [Vietnam gov.]

Post-Doctoral Fellow

Warley Gramacho Da Silva [Tocantins University, Brazil, until Aug 2016]

Visiting Scientists

Tristan Braquelaire [Univ. Bordeaux I, from Mar 2016 until Apr 2016]
Guillaume Chapuis [LANL, Los Alamos, NM, USA, June 2016]
Hristo Djidjev [LANL, Los Alamos, NM, USA, June 2016]

Administrative Assistant

Marie Le Roic [Univ. Rennes I]

2. Overall Objectives

2.1. Optimization of genomic data processing

The first objective of GenScale is the design of scalable, optimized and parallel algorithms for processing the mass of genomic data provided by today biotechnologies. More specifically, our research activities focus on the optimization of the following treatments:

- Processing of HTS data (High Throughput Sequencing) generated by sequencers of 2nd and 3rd generation. These machines generate billions of short DNA fragments (called reads) requiring treatments such as read compression, read correction, genome assembly (contig generation, scaffolding) and detection of variants (Single Nucleotide Polymorphism (SNP), insertion, deletion, inversion, etc.).
- Comparison of large genomic or metagenomic data sets. This fundamental bioinformatics task, due to the steadily increasing of genomic data, is still a bottleneck in many treatments such as taxonomic assignation, functional assignation, genome annotation, etc. Furthermore, the data analysis of large metagenomic projects does not scale with standard sequence comparison methods. New strategies must be investigated.
- 3D protein structure. Functionalities of proteins are mainly supported by their three dimensional structures. Determining these structures from Nuclear Magnetic Resonance (NMR) data or classifying them based on their 3D structures into families require the development of highly optimized algorithms.

Optimization is addressed both in terms of memory space and computation time. Space optimization aims to lower the memory footprint of the algorithms. This is done by the design of innovative data structures. Time optimization aims to provide algorithms with short computation time. Two main ways are followed: combinatorial optimization and multilevel parallelism.

2.2. Active collaboration with life science actors

The second GenScale objective is to create and maintain permanent partnerships with life science research groups. It also aims to be involved in challenging genomic projects in the following areas:

- Health;
- Agronomy and Environment.

GenScale is an interdisciplinary project, which requires strong links with the biology and the genomic scientific community. Hence, it is highly important to keep close relationships with end-users, and being able to have a quick feedback, especially through relevant bioinformatics studies. This is a guarantee for answering right biological questions through right bioinformatics tools.

Collaborations with life science partners go through local, national or international common projects where our tools and methodologies are intensively tested and used. GenScale also welcomes people from INRA, the French research institute in agronomy.

3. Research Program

3.1. Introduction

Based on these overall objectives, the research program of GenScale is structured into four research axes as described below. The first three axes include pure computer science aspects, such as the development of advanced data structures and/or the design of new optimized algorithms; they also include strong partnerships with life science actors to validate the methodologies that are developed. The fourth axis can be seen as a transversal one. It addresses efficient parallel implementations of our methods on standard processors, cluster systems, or accelerators such as GPU.

3.2. Axis 1: HTS data processing

The raw information delivered by NGS (Next Generation Sequencing) technologies represents billions of short DNA fragments. An efficient structuration of this mass of data is the de-Bruijn graph that is used for a large panel of problems dealing with high throughput genomic data processing. The challenge, here, is to represent this graph into memory. An efficient way is to use probabilistic data structures, such as Bloom filters but they generate false positives that introduce noise and may lead to errors. Our approach is to enhance this basic data structure with extra information to provide exact answers, while keeping a minimal memory occupancy [3], [4].

Based on this central data structure, a large panel of HTS algorithms can be designed: read compression, read correction, genome assembly, detection of SNPs (Single Nucleotide Polymorphism) or detection of other variants such as inversion, transposition, etc [10], [8]. The use of this compact structure guarantees software with very low memory footprint that can be executed on many standard-computing resources.

In the full assembly process, an open problem due to the structure complexity of many genomes is the scaffolding step that consists in reordering contigs along the chromosomes. This treatment can be formulated as a combinatorial optimization problem exploiting the upcoming new sequencing technologies based on long reads.

3.3. Axis 2: Sequence comparison

Comparing genomic sequences (DNA, RNA, protein) is a basic bioinformatics task. Powerful heuristics (such as the seed-extend heuristic used in the well-known BLAST software) have been proposed to limit the computation time. The underlying data structures are based on seed indexes allowing a drastic reduction of the search space. However, due to the increasing flux of genomic sequences, this treatment tends to increase and becomes a critical section, especially in metagenomic projects where hundred of millions of reads must be compared to large genomic banks for taxonomic of functional assignation.

Our research follows mainly two directions. The first one revisits the seed-extend heuristic in the context of the bank-to-bank comparison problem. It requires new data structures to better classify the genomic information, and new algorithmic methods to navigate through this mass of data [7], [9]. The second one addresses metagenomic challenges that have to extract relevant knowledge from Tera bytes of data. In that case, the notion of sequence similarity itself is redefined in order to work on objects that are much simpler than the standard alignment score, and that are better suited for large-scale computation. Raw information (reads) is first reduced to k-mers from which high speed and parallel algorithms compute approximate similarities based on a well defined statistical model [5], [2].

3.4. Axis 3: Protein 3D structure

The three-dimensional (3D) structure of proteins tends to be evolutionarily better preserved during evolution than its sequence. Finding structural similarities between proteins gives deep insights into whether these proteins share a common function or whether they are evolutionarily related. Structural similarity between two proteins is usually defined by two functions – a one to- one mapping (also called alignment) between two subchains of their 3D representations and a specific scoring function that assesses the alignment quality. The structural alignment problem is to find the mapping that is optimal with respect to the scoring function. Protein structures can be represented as graphs, and the problem reduces to various combinatorial optimization problems that can be formulated in this framework: for example finding the maximum weighted path [1] or finding the maximum cardinality clique/pseudo-clique [6].

In most cases, however, suitable conformations for a given protein are unknown. To support this statement, we point out that the number of deposited protein conformations on the Protein Data Bank (PDB [0]) recently reached the threshold of 110,000 entries, while the UniProtKB/TrEMBL [0] database contains more than 50 million sequence entries, all of them potentially capable for coding for a new protein. In this context, distance geometry provides powerful methods and algorithms for the identification of protein conformations from Nuclear Magnetic Resonance (NMR) data, which basically consist of a distance list concerning atom pairs of the protein. We are working on the discretization of the distance geometry, so that its search space becomes discrete (and finite!), for making it possible to perform an exhaustive exploration of the solution set.

3.5. Axis 4: Parallelism

Together with the design of new data structures and new algorithms, our research program aims to propose efficient hardware implementation. Even if not explicitly mentioned in the three previous axes, we have constantly in mind to exploit the parallelism of current processors. Practically, and depending on the nature of the computation to perform, three levels of parallelism are addressed: the use of vector instructions of today processors, the multithreading offered by multi-core systems, and the cluster (or cloud) infrastructures.

Consequent bioinformatics treatments, from the processing of raw HTS data to high-level analysis, are generally performed within a workflow environment and executed on cluster systems. Automating the parallelization of such treatments directly from a graphical capture of the workflow is a necessity for end-users that are generally not expert in parallelism. The challenge here is to hide, as much as possible, the different transformations to go from a high level workflow description to an efficient parallel execution that exploits both task-level and data-level parallelism.

Another research activity of this axe is the design of parallel algorithms targeting hardware accelerators, especially GPU boards (Graphical Processing Unit). These devices now offer a high-level programming environment to access the hundred of processors available on a single chip. A few bioinformatics treatments, such as the ones that exhibit good computational regularity, can highly benefit from the computing power of this technology.

4. Application Domains

4.1. Introduction

Today, sequencing data are intensively used in many life science projects. The methodologies developed by the GenScale group are generic approaches that can be applied to a large panel of domains such as health, agronomy or environment areas. The next sections briefly describe examples of our activity in these different domains.

4.2. Health

Cancer diagnostic: from a pool of known genes, the aim is to detect potential mutations that perturb the activity of these genes. Pointing out the right gene helps in prescribing the right drug. The bioinformatics analysis is based on the detection of SNPs (Single Nucleotide Polymorphism) from a set of target genes.

[0] http://www.rcsb.org/
[0] http://www.ebi.ac.uk/uniprot/TrEMBLstats

Microbiology: Streptococcus bacteria are considered as major pathogens for humans and lead to many infections. The cause of their pathogenicity can be studied from their genomic structure by comparing different strains. Text of the genomes must first be constructed (assembly process) before to be analyzed (comparative genomic).

HLA genotyping: The human leukocyte antigen (HLA) system drives the regulation of the human immune system. The HLA genes reside on chromosome 6 and have a large number of alleles. Genotyping this group of genes can be done by a deep sequencing of the HLA region, and by comparing reads with a HLA databank (intensive sequence comparison).

4.3. Agronomy and Environment

Improving plant breeding: such projects aim at 1) identifying favorable alleles at loci contributing to phenotypic variation, 2) characterizing N-traits at the functional level and 3) providing robust multi-locus SNP-based predictors of the breeding value of agronomical traits under polygenic control. Underlying bioinformatics processing is the detection of informative zones (QTL) on the plant genomes.

Insect study: Insects represent major crop pests, justifying the need for control strategies to limit population outbreaks and the dissemination of plant viruses they frequently transmit. Several issues are investigated through the analysis and comparison of their genomes: understanding their phenotypic plasticity such as their reproduction mode changes, identifying the genomic sources of adaptation to their host plant and of ecological speciation, and understanding the relationships with their bacterial symbiotic communities.

Ocean biodiversity: The metagenomic analysis of seawater samples provides an original way to study the ecosystems of the oceans. Through the biodiversity analysis of different ocean spots, many biological questions can be addressed, such as the plankton biodiversity and their role, for example, in the CO2 sequestration.

5. Highlights of the Year

5.1. Highlights of the Year

- **Colib'read Workshop, Nov 7-8 th, Institut Curie, Paris.** GenScale organized a two-day workshop to present the main results of the Colib'read ANR (2013-2016, Coordinator P. Peterlongo) to the scientific community.
- **GATB Programming days.** In 2016, GenScale organized two Genome Analysis Toolbox (GATB) trainings days in Rennes (June 15 th) and Paris (Nov. 9 th). Each event gathered 15 persons who learned how to use the GATB library to design efficient NGS tools.

6. New Software and Platforms

6.1. AskOmics

KEYWORDS: RDF - SPARQL - Querying - Graph
FUNCTIONAL DESCRIPTION

AskOmics allows to load heterogeneous bioinformatics data (formatted as tabular files) into a Triple Store system using a user-friendly web interface. AskOmics also provides an intuitive graph-based user interface supporting the creation of complex queries that currently require hours of manual searches across tens of spreadsheet files. The elements of interest selected in the graph are then automatically converted into a SPARQL query that is executed on the users' data.

- Authors: Charles Bettembourg, Yvanne Chaussin, Anthony Bretaudeau, Olivier Filangi, Fabrice Legeai and Olivier Dameron
- Partners: CNRS - INRA - Université de Rennes 1
- Contact: Fabrice Legeai
- URL: https://github.com/askomics/askomics

6.2. BBhash

Basic binary representation of successive hash
KEYWORDS: C++ - Indexation - Data structures
FUNCTIONAL DESCRIPTION

BBHash is a simple library for building minimal perfect hash function. Given a set of N input keys, it will compute a bijective function that will associate to each key an integer between 1 and N. This then allows to create an indexed array that will hold some data for each key. It is designed to handle large scale datasets (hundred billion and more elements). The function itself is just a little bit larger than other state-of-the-art libraries, it takes approximately 3 bits / elements (compared to 2.62 bits/elem for the emphf lib), but construction is faster and does not require additional memory.

- Participants: Guillaume Rizk, Pierre Peterlongo, Rayan Chikhi and Antoine Limasset
- Contact: Guillaume Rizk
- URL: https://github.com/rizkg/BBHash

6.3. BCALM 2

KEYWORDS: Bioinformatics - NGS - Genomics - Metagenomics - De Bruijn graphs
SCIENTIFIC DESCRIPTION

BCALM 2 is a bioinformatics tool for constructing the compacted de Bruijn graph from sequencing data. It is a parallel algorithm that distributes the input based on a minimizer hashing technique, allowing for good balance of memory usage throughout its execution. It is able to compact very large datasets, such as spruce or pine genome raw reads in less than 2 days and 40 GB of memory on a single machine.
FUNCTIONAL DESCRIPTION

BCALM 2 is an open-source tool for dealing with DNA sequencing data. It constructs a compacted representation of the de Bruijn graph. Such a graph is useful for many types of analyses, i.e. de novo assembly, de novo variant detection, transcriptomics, etc. The software is written in C++ and makes extensive use of the GATB library.

- Participants: Rayan Chikhi, Antoine Limasset and Paul Medvedev
- Contact: Rayan Chikhi
- URL: https://github.com/GATB/bcalm

6.4. BGREAT

De bruijn graph read alignment tool
KEYWORDS: Short reads - Genome assembling
FUNCTIONAL DESCRIPTION

Mapping genomic extracts (reads) on genomic references is a central and necessary task in most genomic studies. But reference sequences are mainly extracted from assembly graphs through an inexact process that both creates chimeras and losses biological pieces of information. This motivates the need of mapping sequences on references represented by graphs. BGREAT is conceived to map reads on de Bruijn graph, a widely used graph in genome assembly.

- Participants: Pierre Peterlongo and Antoine Limasset
- Contact: Pierre Peterlongo
- URL: https://github.com/Malfoy/BGREAT

6.5. GATB-Core

Genome Assembly and Analysis Tool Box
KEYWORDS: Bioinformatics - NGS - Genomics - Genome assembling

Functional Description

The GATB-Core library aims to lighten the design of NGS algorithms. It offers a panel of high-level optimized building blocks to speed-up the development of NGS tools related to genome assembly and/or genome analysis. The underlying data structure is the de Bruijn graph, and the general parallelism model is multithreading. The GATB library targets standard computing resources such as current multicore processor (laptop computer, small server) with a few GB of memory. From high-level API, NGS programming designers can rapidly elaborate their own software based on domain state-of-the-art algorithms and data structures. The GATB-Core library is written in C++.

- Participants: Dominique Lavenier, Guillaume Rizk, Pierre Peterlongo, Charles Deltel, Patrick Durand and Claire Lemaitre
- Contact: Dominique Lavenier
- URL: http://gatb.inria.fr/

6.6. GATB-Core Tutorial

Online GATB-Core tutorial
Keyword: Bioinformatics
Functional Description

"GATB-Core tutorial" is an interactive learning tool that aims at learning software development relying on the bioinformatics toolkit GATB-Core without the need of installing it, its dependencies and a C++ compiler. The tutorial relies on a client-server system. The client is simply a web browser running a full-featured C++ code editor. In turns, it is embedded in templates for the purpose of displaying various lessons. The server side is a Linux-based VM capable of compiling and running "online" any C++ code snippets using GATB-Core. That VM is deployed on Inria's A||Go SaaS platform.

- Participant: Patrick Durand
- Contact: Patrick Durand
- URL: http://gatb-core.gforge.inria.fr/training/

6.7. MindTheGap

Keywords: Bioinformatics - NGS - Genome assembling
Functional Description

MindTheGap performs detection and assembly of DNA insertion variants in NGS read datasets with respect to a reference genome. It is designed to call insertions of any size, whether they are novel or duplicated, homozygous or heterozygous in the donor genome. The main algorithmic improvement of version 2.0.0 is to detect additionnal variants, such as SNPs and deletions. This feature improves the sensitivity of the insertion detection algorithm for insertions that are located near these other variants. Additionnally, MindTheGap performs de novo assembly using the de Bruijn graph implementation of GATB. Hence, the computational resources required to run MindTheGap are significantly lower than that of other assemblers.

- Participants: Claire Lemaitre and Guillaume Rizk
- Contact: Claire Lemaitre
- URL: https://gatb.inria.fr/software/mind-the-gap/

6.8. PLAST

Local alignment tool
Keywords: Bioinformatics - Genomic sequence - Genomics
Functional Description

PLAST is a parallel alignment search tool for comparing large protein banks.

Sequence similarity searching is an important and challenging task in molecular biology and next-generation sequencing should further strengthen the need for faster algorithms to process such huge amount of data. At the same time, the internal architecture of current microprocessors is tending towards more parallelism, leading to the use of chips with two, four and more cores integrated on the same die. The main purpose of this work was to design an effective algorithm to fit with the parallel capabilities of modern microprocessors. A parallel algorithm for comparing large genomic banks and targeting middle-range computers has been developed and implemented in PLAST software. The algorithm exploits two key parallel features of existing and future microprocessors: the SIMD programming model (SSE instruction set) and the multithreading concept (multicore). Compared to multithreaded BLAST software, tests performed on an 8-processor server have shown speedup ranging from 3 to 6 with a similar level of accuracy.

- Participants: Dominique Lavenier, Erwan Drezen and Van Hoa Nguyen
- Contact: Dominique Lavenier
- URL: https://team.inria.fr/genscale/high-throughput-sequence-analysis/plast-intensive-sequence-comparison/

6.9. Simka

KEYWORDS: Comparative metagenomics - K-mer - Distance - Ecology
FUNCTIONAL DESCRIPTION

Simka is a comparative metagenomics method dedicated to NGS datasets. It computes a large collection of distances classically used in ecology to compare communities by approximating species counts by k-mer counts. The method scales to a large number of datasets thanks to an efficient and parallel kmer-counting strategy that processes all datasets simultaneoulsy.

- Participants: Gaetan Benoit, Claire Lemaitre, Pierre Peterlongo and Dominique Lavenier
- Contact: Gaetan Benoit
- URL: https://gatb.inria.fr/software/simka/

6.10. short read connector

KEYWORDS: Bioinformatics - Genomics - Metagenomics
SCIENTIFIC DESCRIPTION

Short read connector enables the comparisons of two read sets B and Q. For each read from Q it provides either:

The number of occurrences of each k-mers of the read in the set B (SRC_counter) or A list of reads from B that share enough k-mers with the tested read from B (SRC_linker)
FUNCTIONAL DESCRIPTION

This tool uses a data structure (BBHASH) adapted to the indexing of big data. Short Read Connector works on reads, which are sequencing data from high-throughput sequencers. Once the data is indexed, short read connector makes it possible either to find the similar reads in a dataset or to simply retrieve the approximate number of these similar reads.

- Participants: Pierre Peterlongo, Camille Marchet and Antoine Limasset
- Partner: UPMC
- Contact: Pierre Peterlongo
- URL: https://github.com/pierrepeterlongo/short_read_connector

7. New Results

7.1. HTS data processing

7.1.1. Providing end-user solutions, example from the Colib'read on galaxy project

Participants: Claire Lemaitre, Camille Marchet, Pierre Peterlongo.

Colib'read tools suite uses optimized reference-free algorithms for various analyses of NGS datasets, such as variant calling or read set comparisons. To facilitate data analysis and tools dissemination, we developed Galaxy tools and tool shed repositories. The galaxy package, facilitates the analyzis of raw NGS data for a broad range of life scientists [16].

7.1.2. Assembly of Streptococcus Bacteria

Participant: Dominique Lavenier.

With the microbiological and bacteriological group of the Rennes hospital, we design a new strategy to assemble the genomes of 40 *Streptococcus* bacteria. Each strain has been sequenced and independently assembled using different assembly tools. For a specific strain, a merge of the contigs is done using the MIX software. This step allows the number of contigs to be significantly reduced, resulting in a better final assembly compared to each individual assembly. The comparison with other known *Streptococcus* genomes indicates where phages are located in the genome [20].

7.1.3. Data-mining applied to GWAS

Participants: Pham Hoang Sun, Dominique Lavenier.

Identifying variant combination association with disease is a bioinformatics challenge. This problem can be solved by discriminative pattern mining that uses statistical functions to evaluate the significance of individual biological patterns. There is a wide range of such measures. However, selecting an appropriate measure as well as a suitable threshold in some specific practical situations is a difficult task. We propose to use the skypattern technique which allows combinations of measures to be used to evaluate the importance of variant combinations without having to select a given measure and a fixed threshold. Experiments on several real variant datasets demonstrate that the skypattern method effectively identifies the risk variant combinations related to diseases [28].

7.1.4. Variant detection in transcriptomic data

Participant: Camille Marchet.

We defined a method to identify, quantify and annotate SNPs (Single Nucleotide Polymorphisms) using RNA-seq reads only. Organisms with a poor quality or no reference genome can take benefit of this approach, as well as studies where not enough material is available for sequencing from one individual, where samples can be pooled. The method relies on motifs discovery and post-treatment in de Bruijn graphs built from the reads. It can be used for any species to annotate SNPs and predict their impact on proteins as well as test their association to a phenotype of interest. The approach has been validated using well known human RNA-seq data. Results have been compared with state of the art approaches for variant calling. We showed that the methods perform similarly in terms of precision and recall. Then we focused on the main target of the study, namely the non-model species. We finally validated experimentally the predictions of our method [18].

7.1.5. Faster de Bruijn graph compaction

Participant: Antoine Limasset.

We developed a new algorithm, called BCALM2, for the compaction of de Bruijn graphs. BCALM2 is a parallel algorithm based on minimizer repartition of sequences. This repartition allows the compaction of extremely large graphs with moderate memory usage and time. The compaction of a human sequencing graph can be done in 1 hour with only 3GB of memory and huge genomes, such as the pine and white spruce ones (more than 20Gbp each), can be handled using our approach on a regular server (2 days and 40GB of memory). Those results argue that BCALM2 is one order of magnitude more efficient than available approaches and can tackle the assembly bottleneck of constructing a compacted de Bruijn graph [14].

7.1.6. Scaffolding

Participants: Rumen Andonov, Sebastien François, Dominique Lavenier.

We developed a method for solving genome scaffolding as a problem of finding a long simple path in a graph defined by the contigs that satisfies additional constraints encoding the insert-size information. Then we solved the resulting mixed integer linear program to optimality using the Gurobi solver. We tested our algorithm on several chloroplast genomes and showed that it outperforms other widely-used assembly solvers by the accuracy of the results [25].

7.2. Sequence comparison

7.2.1. Metagenomics datasets comparison

Participants: Gaetan Benoit, Dominique Lavenier, Claire Lemaitre, Pierre Peterlongo.

We developed a new method, called Simka, to compare simultaneously numerous large metagenomics datasets. The method computes pairwise distances based on the amount of shared k-mers between datasets. The method scales to a large number of datasets thanks to an efficient kmer-counting step that processes all datasets simultaneoulsy. Additionnally, several distance definitions were implemented and compared, including some originating from the ecological domain. The method is currently applied to the TARA oceans project (more than 2000 datasets) which aims at comparing worldwide sea water samples (ANR HydrGen project) [12].

7.2.2. Read similarity detection

Participants: Camille Marchet, Antoine Limasset, Pierre Peterlongo.

Retrieving similar reads inside or between read sets is a fundamental task either for algorithmic reasons or for analyses of biological data. This task is easy in small datasets, but becomes particularly hard when applied to millions or billions of reads. In [24] we used a straightforward indexing structure that scales to billions of elements. We proposed two direct applications in genomics and metagenomics. These applications consist in either approximating the number of similar reads between dataset(s) or to simply retrieve these similar reads. They can be applied on distinct read sets or on a read set against itself.

7.3. Parallelism

7.3.1. Processing-in-Memory

Participants: Charles Deltel, Dominique Lavenier.

The concept of PIM (Processor In Memory) aims to dispatch the computer power near the data. Together with the UPMEM company (http://www.upmem.com/), which is currently developing a DRAM memory enhanced with computing units, we investigate the parallelization of two bioinformatics algorithms for this new type of memory: sequence alignment and mapping [34] [33]. The first results show that blast-like algorithms or mapping algorithms can highly benefit from such memory and speed-up of more than 25 can be achieved [26].

7.3.2. GPU for graph algorithms

Participants: Rumen Andonov, Dominique Lavenier.

We describe three algorithms and their associated GPU implementations for two types of shortest path problems. These implementations target computations on graphs with up to millions of vertices and executions on GPU clusters. The first two algorithms solve the All-Pairs Shortest Path (APSP) problem. The first of these two algorithms allows computations on graphs with negative edges while the second trades this ability for better parallel scaling properties and improved memory access. The third algorithm solves the Single-Pair Shortest Path (SPSP) query problem. Our implementations efficiently exploit the computational power of 256 GPUs simultaneously. All shortest paths of a million vertex graph can be computed in 6 minutes and shortest path queries on the same graph are answered in a quarter of a millisecond. These implementations proved to be orders of magnitude faster than existing parallel approaches[30].

7.4. Data representation

7.4.1. Computational pan-genomics: status, promises and challenges

Participant: Pierre Peterlongo.

We took part to the Computational Pan-Genomics Consortium producing a "white paper" dedicated to computational pan-genomic. A pan-genome is a representation of the union of the genomes of closely related individuals (eg from a same species). Computational pan-genomics is a new sub-area of research in computational biology. In [19], we generalized existing definitions and we examined already available approaches to construct and use pan-genomes, discussed the potential benefits of future technologies and methodologies and reviewed open challenges from the vantage point of the above-mentioned biological disciplines.

7.4.2. Mapping reads on graphs

Participants: Pierre Peterlongo, Antoine Limasset.

Many published genome sequences remain in the state of a large set of contigs. Each contig describes the sequence found along some path of the assembly graph, however, the set of contigs does not record all the sequence information contained in that graph. Although many subsequent analyses can be performed with the set of contigs, one may ask whether mapping reads on the contigs is as informative as mapping them on the paths of the assembly graph.

In [17], we proposed a formal definition of mapping a sequence on a de Bruijn graph, we analysed the problem complexity, and we provided a practical solution. The proposed tool can map millions of reads per CPU hour on a de Bruijn graph built from a large set of human genomic reads. Results show that up to 22 % more reads can be mapped on the graph but not on the contig set.

7.5. Applications

7.5.1. Study of the rapeseed genome structure

Participants: Sebastien Letort, Pierre Peterlongo, Dominique Lavenier, Claire Lemaitre, Fabrice Legeai.

In collaboration with IGEPP (Institut de Génétique, Environnement et Protection des Plantes), INRA, and through two national projects, PIA Rapsodyn and France-Génomique Polysuccess, we are involved in the genome analysis of several rapeseed varieties. The Rapsodyn project has the ambition to insure long-term competitiveness of the rapeseed production through improvement of the oil yield and reduction of nitrogen inputs during the crop cycle. Rapeseed varieties must thus be selected from genotypes that favor low nitrogen input. DiscoSnp++ is here used to locate new variants among the large panel of rapeseed varieties which have been sequenced during the project.

The PolySuccess project aims to answer the following question: how a polyploid, such as the oilseed rape plant, becomes a new species? Oilseed rape (Brassica napus) being a natural hybrid between B.rapa and B.oleracea, different genomes of these three species have been sequenced to study their structures. The Minia assembly pipeline provides a fast way to generate contigs that are used for studying gene specificities.

7.5.2. GATB Production Pipeline

Participants: Patrick Durand, Charles Deltel.

The entire set of libraries and tools related to the GATB Software have been introduced within a professional environment to support high-quality C++ developments. It relies on the use of technology platforms available at Inria: OpenStack and Jenkins. Considering the latter, we have setup more than 50 Jenkins tasks to automate the entire software development based on GATB: C++ code compiling and testing, documentation creation, packaging and preparation of official releases, mirroring on public Github repositories. Code compilation and tests are done on Linux and MacOSX VMs. https://ci.inria.fr/gatb-core/

7.5.3. Variant predictions in the pea genome

Participant: Pierre Peterlongo.

Progress in genetics and breeding in pea suffered from the limited availability of molecular resources. SNP markers that can be identified through affordable sequencing processes without the need for prior genome reduction or a reference genome allow the discovery of thousands of molecular markers.

We have been involved with IGEPP (Institut de Génétique, Environnement et Protection des Plantes, INRA) in the application of the discoSnp++ tool, discovering SNPs on HiSeq whole genome sequencing of four pea lines. Validation of a subset of predicted SNPs showed that almost all generated SNPs are highly designable and that most (95 %) deliver highly qualitative genotyping result [13].

7.5.4. Analysis of insect pest genomes

Participant: Fabrice Legeai.

Within a large international network of biologists, GenScale has contributed to various projects for identifying important components involved in the adaptation of major agricultural pests to their environment. We provided the assemblies, the annotations and the comparisons of various insects genomes [29]. Following specific agreement or policy, these results are available for browsing and consulting to a restricted consortium or a large community through the BioInformatics platform for Agro-ecosystems Arthropods (http://bipaa.genouest.org/is). In particular, this year our work helped to identify aphid genes involved in the adaptation to their favorite plant [15], or genes that are differentially expressed between leaf- and root-feeding phylloxera [21]. Furthermore, in order to help scientists to consult and cross genomics and postgenomics data, we are developping AskOmics, an integration and interrogation software for (linked) biological data, within a strong partnership, with Dyliss and GenOuest [36], [27].

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Contracts with Industry

8.1.1. Empowerd memory

Participants: Charles Deltel, Dominique Lavenier.

The UPMEM company is currently developing new memory devices with embedded computing power (http://www.upmem.com/). GenScale investigates how bioinformatics algorithms can benefit from these new types of memory (see section New Results).

8.2. Bilateral Grants with Industry

8.2.1. EnginesOn start-up project

Participant: Jennifer Del Giudice.

EnginesOn is a start-up project based on life science digital data analysis (http://engineson.fr/). The origin of the project comes from a simple field observation: NGS technology is involved in numerous scientific studies. Deciphering the heterogeneous and voluminous data generated is a real challenge. People with the skills to analyze this type of data are scarce. EnginesOn focuses its first effort on health market with cancer diagnosis and personalized medecine. The start-up provides to physicians a virtual research laboratory with analysis workflows, compute infrastructure and data management that will lead to a simple, fast, reproducible diagnosis in a transparent fashion. EnginesOn also addresses the issue of big data management and storage. The project is entitled to the Fasttrack program since october 2016. Inria funds a 6-month technology transfer engineer in order to study the valorization and promote the GATB toolbox.

8.2.2. Rapsodyn project

Participants: Dominique Lavenier, Claire Lemaitre, Sebastien Letort, Pierre Peterlongo.

RAPSODYN is a long term project funded by the IA French program (Investissement d'Avenir) and several field seed companies, such as Biogemma, Limagrain and Euralis (http://www.rapsodyn.fr/). The objective is the optimization of the rapeseed oil content and yield under low nitrogen input. GenScale is involved in the bioinformatics work package, in collaboration with Biogemma's bioinformatics team, to elaborate advanced tools dedicated to polymorphism.

9. Partnerships and Cooperations

9.1. Regional Initiatives

9.1.1. Rennes Hospital, Hematology service, Genetic service

Participants: Patrick Durand, Dominique Lavenier, Claire Lemaitre, Pierre Peterlongo, Guillaume Rizk.

The collaboration with the Hematology service and with the Genetic service of the Rennes hospital aims to set up advanced bioinformatics pipelines for cancer diagnosis. More precisely, we are in the process of setting up and evaluating a new method of predictions of small cancer-related mutations (such as SNPs and small insertions/deletions) from raw DNA sequencing data. The method relies on the use of k-mers and clustering of reads to call for mutations. Current prototype relies on Python programming language just for the purpose of evaluating the prediction quality of the software. However, final software is expected to use GATB library to highly increase the performance of the new tool.

9.1.2. Partnership with INRA in Rennes

Participants: Cervin Guyomar, Dominique Lavenier, Fabrice Legeai, Claire Lemaitre, Sébastien Letort, Pierre Peterlongo.

The GenScale team has a strong and long term collaboration with biologists of INRA in Rennes: IGEPP and PEGASE units. This partnership concerns both service and research activities and is acted by the hosting of one INRA engineer (F. Legeai) and one PhD student (C. Guyomar).

9.2. National Initiatives

9.2.1. ANR

9.2.1.1. Project ADA-SPODO: Genetic variation of Spodoptera Frugiperda

Participants: Claire Lemaitre, Fabrice Legeai, Anaïs Gouin, Dominique Lavenier, Pierre Peterlongo.

Coordinator: E. D'Alençon (Inra, Montpellier)
Duration: 45 months (Oct. 2012 – May 2016)
Partners: DGIMI Inra Montpellier, CBGP Inra Montpellier, URGI Inra Versailles, Genscale Inria/IRISA Rennes.

The ADA-SPODO project aims at identifying all sources of genetic variation between two strains of an insect pest: Lepidoptera Spodoptera Frugiperda in order to correlate them with host-plant adaptation and speciation. GenScale's task is to develop new efficient methods to compare complete genomes along with their postgenomic and regulatory data.

9.2.1.2. Project COLIB'READ: Advanced algorithms for NGS data

Participants: Pierre Peterlongo, Antoine Limasset, Camille Marchet, Claire Lemaitre, Dominique Lavenier, Fabrice Legeai, Guillaume Rizk, Chloé Riou.

Coordinator: P. Peterlongo (Inria, GenScale, Rennes)
Duration: 45 months (Mar. 2013 – Dec. 2016)

Partners: LIRMM Montpellier, Erable Inria Lyon, Genscale Inria/IRISA Rennes.

The main goal of the Colib'Read project is to design new algorithms dedicated to the extraction of biological knowledge from raw data produced by High Throughput Sequencers (HTS). The project proposes an original way of extracting information from such data. The goal is to avoid the assembly step that often leads to a significant loss of information, or generates chimerical results due to complex heuristics. Instead, the strategy proposes a set of innovative approaches that bypass the assembly phase, and that do not require the availability of a reference genome. https://colibread.inria.fr/

9.2.1.3. Project HydroGen: Metagenomic applied to ocean life study

Participants: Dominique Lavenier, Pierre Peterlongo, Claire Lemaitre, Guillaume Rizk, Gaëtan Benoit.

Coordinator: P. Peterlongo (Inria/Irisa, GenScale, Rennes)
Duration: 42 months (Nov. 2014 – Apr. 2018)
Partners: CEA (GenosScope, Evry), INRA (AgroParisTech, Paris – MIG, Jouy-en-Jossas).

The HydroGen project aims to design new statistical and computational tools to measure and analyze biodiversity through comparative metagenomic approaches. The support application is the study of ocean biodiversity based on the analysis of seawater samples available from the Tara Oceans expedition.

9.2.1.4. Project SpeCrep: speciation processes in butterflies

Participants: Dominique Lavenier, Pierre Peterlongo, Claire Lemaitre, Fabrice Legeai.

Coordinator: M. Elias (Museum National d'Histoire Naturelle, Institut de Systematique et d'Evolution de la Biodiversite, Paris)
Duration: 48 months (Jan. 2015 – Dec. 2018)
Partners: MNHN (Paris), INRA (Versailles-Grignon), Genscale Inria/IRISA Rennes.

The SpeCrep project aims at better understanding the speciation processes, in particular by comparing natural replicates from several butterfly species in a suture zone system. GenScale's task is to develop new efficient methods for the assembly of reference genomes and the evaluation of the genetic diversity in several butterfly populations.

9.2.2. PIA: Programme Investissement d'Avenir

9.2.2.1. RAPSODYN: Optimization of the rapeseed oil content under low nitrogen

Participants: Dominique Lavenier, Claire Lemaitre, Sebastien Letort, Pierre Peterlongo.

Coordinator: N. Nesi (Inra, IGEPP, Rennes)
The objective of the Rapsodyn project is the optimization of the rapeseed oil content and yield under low nitrogen input. GenScale is involved in the bioinformatics work package to elaborate advanced tools dedicated to polymorphism and application to the rapeseed plant.

9.2.2.2. France Génomique: Bio-informatics and Genomic Analysis

Participants: Laurent Bouri, Dominique Lavenier.

Coordinator: J. Weissenbach (Genoscope, Evry)
France Génomique gathers resources from the main French platforms in genomic and bio-informatics. It offers to the scientific community an access to these resources, a high level of expertise and the possibilities to participate in ambitious national and international projects. The GenScale team is involved in the work package "assembly" to provide expertise and to design new assembly tools for the 3rd generation sequencing.

9.2.3. Programs from research institutions

9.2.3.1. Inria ADT DiagCancer

Participants: Dominique Lavenier, Patrick Durand.

Since October 1st, 2016, Genscale has started a one-year Inria ADT called DiagCancer. It aims at: (1) including the DiscoSnp++ tool within the current data production pipeline at Pontchaillou Hospital (Rennes), (2) providing a new prediction tool applied to the calling of cancer related mutations from DNA sequencing data and (3) creating new analysis tools to facilitate the interpretation of results by end-users (biologists, doctors). The project is done in close collaboration with Haematology Service, CHU Pontchaillou, Rennes.

9.3. International Initiatives

9.3.1. Informal International Partners

- Free University of Brussels, Belgium: Genome assembly [P. Perterlongo, R. Andonov]
- IMECC, UNICAMP, Campinas, Brazil: Distance geometry problem [A. Mucherino]
- Los Alamos National Laboratory (LANL), Los Alamos: Graph structure, Parallelism, GPU [R. Andonov, D. Lavenier, G. Rizk]

9.4. International Research Visitors

9.4.1. Visits of International Scientists

- Visit of prof. Tomi Klein from Bar Ilan University (Israel). One week, december 2016. Collaboration for the application of approximate hash function to the TGS data analysis [P. Peterlongo]
- Visit of Hristo Djidjev from Los Alamos National Laboratory, June 2016. Graph algorithms for scaffolding problem, professeur invité, University of Rennnes 1, [R. Andonov]
- Visit of Guillaume Chapuis from Los Alamos National Laboratory, June 2016. Parallelism, GPU. [R. Andonov, D. Lavenier]

9.4.2. Internships

- Samyadeep Basu, BITS Pilani, India, May - July 2016. Development of a web server for assembling bacteria genomes [D. Lavenier, P. Durand, C. Deltel]

9.4.3. Visits to International Teams

9.4.3.1. Research Stays Abroad

- Visit of Guillaume Rizk to Los Alamos National Laboratory, USA, August - September 2016 (2 months). Efficient combinatorial optimization using quantum computing.

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events Organisation

10.1.1.1. General Chair, Scientific Chair

- Workshop Colib'read. Scientific and practical organization [C. Lemaitre, C. Marchet, P. Peterlongo]

10.1.2. Scientific Events Selection

10.1.2.1. Chair of Conference Program Committees

- WCO16, Gdansk, Poland (co-chair) [A. Mucherino]

10.1.2.2. Member of the Conference Program Committees

- BIBM 2016: IEEE International Conference on Bioinformatics and Biomedicine [D. Lavenier]
- BIOKDD 2016: 15th International Workshop on Data Mining in Bioinformatics [D. Lavenier]
- ACM PASC16 Conference : Platform for Advanced Scientific Computing [D. Lavenier]
- RECOMB 2016 : International Conference on Research in Computational Molecular Biology [D. Lavenier]
- ECCB 2016 : 15th European Conference on Computational Biology [P. Peterlongo]
- SeqBio 2016 : Bioinformatics multidisciplinary workshop [P. Peterlongo]
- WACEBI 2016 : Workshop on Accelerator-Enabled Algorithms and Applications in Bioinformatics [D. Lavenier]

10.1.2.3. Reviewer

- RECOMB 2017 [R. Andonov, C. Lemaitre]
- ECCB 2016 [P. Peterlongo]

10.1.3. Journal

10.1.3.1. Reviewer - Reviewing Activities

- Algorithms for Molecular Biology [P. Peterlongo]
- Bioinformatics [D. Lavenier, P. Peterlongo]
- BMC Bioinformatics [P. Peterlongo]
- Drug Dicovery Today [D. Lavenier]
- Journal of Parallel and Distributed Computing [D. Lavenier]
- Plos One [D. Lavenier, R. Andonov]
- Journal of Biomedical and Health Informatics [D. Lavenier]
- Briefing in Bioinformatics [D. Lavenier]
- Theoretical Computer Science [P. Peterlongo]
- Fundamenta Informaticae, Integrated Computer-Aided Engineering (IOS Press) [A. Mucherino]
- COMPAG, Elsevier [A. Mucherino]

10.1.4. Invited Talks

- D. Lavenier, *Parallel Processing of Sequencing Data*, Conférence d'informatique en Parallélisme, Architecture et Système, Lorient, France, July 2016
- D. Lavenier *GATB: A Genomic Analysis Tool Box for designing parallel and low memory fingerprint bioinformatics software*, Pasteur Institut, Paris, France, Sep. 2016.
- D. Lavenier, *Low memory fingerprint data structure for genomics*, ENS Rennes, France, oct. 2016.
- D. Lavenier *GATB: A Genome Analysis Tool Box for designing parallel and low memory footprint bioinformatics software*, Workshop on Emerging Bioinformatics Applications for Microbial Ecogenomics, Brest, France, oct. 2016.
- C. Lemaitre, *Comparing numerous metagenomics datasets*, Laboratoire de Biométrie et Biologie Évolutive, Lyon, France, Nov. 2016.
- P. Peterlongo *De novo comparison of (large number of) metagenomic samples*, Metagenomics day, Billile, Lille, France June 2016.
- P. Peterlongo *Finding SNPs de novo from reads* , Bioadvection workshop, Napoli, Italy, June 2016.
- P. Peterlongo *De novo comparison of (large number of) metagenomic samples, What are the technical challenges? what can we expect from this?*, RCAM workshop - keynote speaker, The Hague, Netherlands, September 2016.

- P. Peterlongo *Multiple Comparative Metagenomics using Multiset k-mer Counting*, Pasteur Summer School 2016 In Metagenomics, September 2016.
- C. Lemaitre *Comparaison (massive) de (nombreux) metagénomes. Passons par les kmers pour passer à l'échelle*, Journée scientifique sur "le Microbiome" organisée par Biogenouest, Rennes, France, December 2016.
- F. Legeai *Les analyses bioinformatiques pour les données épigénomiques*, Atelier ChIP du réseau REacTION, Paris, France, December 2016.
- R. Andonov, *Global Optimization Methods for Genome Scaffolding and Completing Genome Assemblies*, Workshop on Graph Assembly Algorithms for omics data, Univ. Milano-Bicocca, Italy, November 18, 2016

10.1.5. Leadership within the Scientific Community

- P. Peterlongo. Animator of one of the scientific axes of the GDR BIM group of research.
- P. Peterlongo. Member of the SFBI board.

10.1.6. Scientific Expertise

- Expert for the MEI (International Expertise Mission), French Research Ministry [D. Lavenier]
- Member of the Scientific Council of BioGenOuest [D. Lavenier]
- Member of the Scientific Council of the Computational Biology Institute of Montpellier [D. Lavenier]

10.1.7. Research and Pedagogical Administration

- Member of the CoNRS, section 06, [D. Lavenier]
- Member of the local Inria Rennes CDT (Technologic Transfer Commission) [D. Lavenier]
- Member of the steering committee of the INRA BIPAA Platform (BioInformatics Platform for Agro-ecosystems Arthropods) [D. Lavenier]
- Member of the steering committee of The GenOuest Platform (Bioinformatics Platform of BioGenOuest) [D. Lavenier]
- Representative of the environnemental axis of UMR IRISA [C. Lemaitre]
- AGOS first secretary [P. Peterlongo]
- Organisation of the weekly seminar "Symbiose" [P. Peterlongo]
- Scientific Responsible for International Relationships at ISTIC [A. Mucherino]
- Member of "Commission Affaires Internationales" at University of Rennes 1 [A. Mucherino]
- In charge of the bachelor's degree in the computer science department of University of Rennes 1 (90 students) [R. Andonov]

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

Licence : R. Andonov, Graph Algorithms, 60h, Univ. Rennes 1, France.

Licence : A. Mucherino, Java basis, 80h, L1, Univ. Rennes 1, France.

Master : A. Mucherino, R. Andonov, Operational Research, 78h, M1, Univ. Rennes 1, France.

Master : A. Mucherino, Introduction to Computational Systems and Networks, 42h, M1, Univ. Rennes 1, France.

Master : A. Mucherino, Object Oriented Programming, 40h, M1, Univ. Rennes 1, France.

Master : A. Mucherino, P. Peterlongo and R. Andonov, Algorithms on Sequences and Structures, 36h, M2, Univ. Rennes 1, France.

Master : A. Mucherino, Parallel Computing (in English), 18h, M1, Univ. Rennes 1, France.

Master : C. Lemaitre, Dynamical systems for biological networks, 20h, M2, Univ. Rennes 1, France.

Master : C. Lemaitre, P. Peterlongo, Algorithms on Sequences for Bioinformatics, 24h, M1, Univ. Rennes 1, France.

Master : P. Peterlongo, Experimental Bioinformactics, 12h, M1, ENS Rennes, France.

Master : D. Lavenier, Research training module, 24h, M1, Univ. Rennes 1, Rennes, France.

Master : R. Andonov, Advanced Algorithmics, 25h, Univ. Rennes 1, France.

Training : P. Durand, G. Rizk, GATB Programming Day, 8h (June 16th), Univ. Rennes 1, Rennes, France.

Training : P. Durand, G. Rizk, GATB Programming Day, 8h (May 9th), Institut Henri Poincaré, France.

E-learning

Online tutorial : GATB-Core Online Training tool. http://gatb-core.gforge.inria.fr/training/

discoSnp tutorial available from the discoSnp webpage: https://colibread.inria.fr/software/discosnp/

10.2.2. Supervision

HdR : Pierre Peterlongo, Lire les lectures : analyse de données de séquençage, Univ. Rennes 1, 25/01/2016, [11] https://hal.inria.fr/tel-01278275.

PhD in progress : G. Benoit, New algorithms for comparative metagenomics, 01/11/2014, D. Lavenier and C. Lemaitre.

PhD in progress : A. Limasset, Algorithm for Genomics, 09/2014, D. Lavenier and P. Peterlongo.

PhD in progress : C. Guyomar, Bioinformatic tools and applications for metagenomics of bacterial communities associated to insects, 01/10/2015, C. Lemaitre and F. Legeai.

PhD in progress : C. Marchet, Nouvelles méthodologies pour l'assemblage de données de séquençage polymorphes, 01/10/2015, P. Peterlongo.

PhD in progress : P. Hoan Son, Data mining and bioinformatics, 01/2015, D. Lavenier and A. Termier.

PhD in progress : S. François, Combinatorial Optimization Approaches for Bioinformatics, 01/10/2016, R. Andonov.

10.2.3. Juries

- *President of Ph-D thesis jury*. Phuong Do Viet, University of Montpellier [R. Andonov], Karel Brinda, University of Marne la Vallée [D. Lavenier]
- *Member of Ph-D thesis juries*. Joseph Lucas, University Pierre et Marie Curie [C. Lemaitre].
- *Referee of Ph-D thesis*. Cécile Monat, University of Montpellier [D. Lavenier], C. Vroland, University of Lille [P. Peterlongo].
- *Member of Ph-D thesis comitees*. L. Ishi Soares de Lima, University of Lyon [C. Lemaitre], Yoann Aigu, University of Rennes [F. Legeai], Hélène Boulain, University of Rennes [F. Legeai], Chunxiang Hao, University of Rennes [D. Lavenier], Alix Mas, University of Rennes [P. Peterlongo], Pierre Charrier, University of Nantes [P. Peterlongo], Cervin Guyomar, University of Rennes [P. Peterlongo], Lea Siegwald, University of Lille [P. Peterlongo].

10.3. Popularization

- Operation "A la découverte de la recherche" [P. Peterlongo]

11. Bibliography

Major publications by the team in recent years

[1] R. ANDONOV, N. MALOD-DOGNIN, N. YANEV.*Maximum Contact Map Overlap Revisited*, in "Journal of Computational Biology", January 2011, vol. 18, n^{o} 1, p. 1-15 [*DOI :* 10.1089/CMB.2009.0196], http://hal.inria.fr/inria-00536624/en.

[2] G. BENOIT, P. PETERLONGO, M. MARIADASSOU, E. DREZEN, S. SCHBATH, D. LAVENIER, C. LEMAITRE.*Multiple comparative metagenomics using multiset k -mer counting*, in "PeerJ Computer Science", November 2016, vol. 2 [*DOI :* 10.7717/PEERJ-CS.94], https://hal.inria.fr/hal-01397150.

[3] R. CHIKHI, G. RIZK.*Space-efficient and exact de Bruijn graph representation based on a Bloom filter*, in "Algorithms for Molecular Biology", 2013, vol. 8, n^{o} 1, 22 [*DOI :* 10.1186/1748-7188-8-22], http://hal.inria.fr/hal-00868805.

[4] E. DREZEN, G. RIZK, R. CHIKHI, C. DELTEL, C. LEMAITRE, P. PETERLONGO, D. LAVENIER.*GATB: Genome Assembly & Analysis Tool Box*, in "Bioinformatics", 2014, vol. 30, p. 2959 - 2961 [*DOI :* 10.1093/BIOINFORMATICS/BTU406], https://hal.archives-ouvertes.fr/hal-01088571.

[5] N. MAILLET, C. LEMAITRE, R. CHIKHI, D. LAVENIER, P. PETERLONGO.*Compareads: comparing huge metagenomic experiments*, in "RECOMB Comparative Genomics 2012", Niterói, Brazil, October 2012, https://hal.inria.fr/hal-00720951.

[6] N. MALOD-DOGNIN, R. ANDONOV, N. YANEV.*Maximum Cliques in Protein Structure Comparison*, in "SEA 2010 9th International Symposium on Experimental Algorithms", Naples, Italy, P. FESTA (editor), Springer, May 2010, vol. 6049, p. 106-117 [*DOI :* 10.1007/978-3-642-13193-6_10], https://hal.inria.fr/inria-00536700.

[7] V. H. NGUYEN, D. LAVENIER.*PLAST: parallel local alignment search tool for database comparison*, in "Bmc Bioinformatics", October 2009, vol. 10, 24, http://hal.inria.fr/inria-00425301.

[8] G. RIZK, A. GOUIN, R. CHIKHI, C. LEMAITRE.*MindTheGap: integrated detection and assembly of short and long insertions*, in "Bioinformatics", December 2014, vol. 30, n^{o} 24, p. 3451 - 3457 [*DOI :* 10.1093/BIOINFORMATICS/BTU545], https://hal.inria.fr/hal-01081089.

[9] G. RIZK, D. LAVENIER.*GASSST: Global Alignment Short Sequence Search Tool*, in "Bioinformatics", August 2010, vol. 26, n^{o} 20, p. 2534-2540, http://hal.archives-ouvertes.fr/hal-00531499.

[10] R. URICARU, G. RIZK, V. LACROIX, E. QUILLERY, O. PLANTARD, R. CHIKHI, C. LEMAITRE, P. PETERLONGO.*Reference-free detection of isolated SNPs*, in "Nucleic Acids Research", November 2014, p. 1 - 12 [*DOI :* 10.1093/NAR/GKU1187], https://hal.inria.fr/hal-01083715.

Publications of the year

Doctoral Dissertations and Habilitation Theses

[11] P. PETERLONGO.*Lire les lectures : analyse de données de séquençage*, Université rennes1, January 2016, Habilitation à diriger des recherches, https://hal.inria.fr/tel-01278275.

Articles in International Peer-Reviewed Journal

[12] G. Benoit, P. Peterlongo, M. Mariadassou, E. Drezen, S. Schbath, D. Lavenier, C. Lemaitre.*Multiple comparative metagenomics using multiset k -mer counting*, in "PeerJ Computer Science", November 2016, vol. 2 [*DOI :* 10.7717/peerj-cs.94], https://hal.inria.fr/hal-01397150.

[13] G. Boutet, S. Alves Carvalho, M. Falque, P. Peterlongo, E. Lhuillier, O. Bouchez, C. Lavaud, M.-L. Pilet-Nayel, N. Rivière, A. Baranger.*SNP discovery and genetic mapping using genotyping by sequencing of whole genome genomic DNA from a pea RIL population*, in "BMC Genomics", December 2016, vol. 17, n^{o} 1, 121 [*DOI :* 10.1186/s12864-016-2447-2], https://hal.inria.fr/hal-01275696.

[14] R. Chikhi, A. Limasset, P. Medvedev.*Compacting de Bruijn graphs from sequencing data quickly and in low memory*, in "Bioinformatics", November 2016, vol. 32, n^{o} 12, i201 - i208 [*DOI :* 10.1093/bioinformatics/btw279], https://hal.archives-ouvertes.fr/hal-01395704.

[15] I. Eyres, J. Jaquiéry, A. Sugio, L. Duvaux, K. Gharbi, J.-J. Zhou, F. Legeai, M. Nelson, J.-C. Simon, C. M. Smadja, R. Butlin, J. Ferrari.*Differential gene expression according to race and host plant in the pea aphid*, in "Molecular Ecology", 2016, vol. 25, n^{o} 17, p. 4197-4215 [*DOI :* 10.1111/mec.13771], https://hal-univ-rennes1.archives-ouvertes.fr/hal-01371828.

[16] Y. Le Bras, O. Collin, C. MONJEAUD, V. Lacroix, E. Rivals, C. Lemaitre, V. Miele, G. Sacomoto, C. Marchet, B. Cazaux, A. Zine El Aabidine, L. Salmela, S. Alves-Carvalho, A. Andrieux, R. Uricaru, P. Peterlongo.*Colib'read on galaxy: a tools suite dedicated to biological information extraction from raw NGS reads*, in "GigaScience", February 2016, vol. 5, n^{o} 1 [*DOI :* 10.1186/s13742-015-0105-2], https://hal.inria.fr/hal-01280238.

[17] A. Limasset, B. Cazaux, E. Rivals, P. Peterlongo.*Read mapping on de Bruijn graphs*, in "BMC Bioinformatics", December 2016, vol. 17, n^{o} 1 [*DOI :* 10.1186/s12859-016-1103-9], https://hal.inria.fr/hal-01349636.

[18] H. Lopez-Maestre, L. Brinza, C. Marchet, J. Kielbassa, S. Bastien, M. Boutigny, D. Monnin, A. El Filali, C. M. Carareto, C. Vieira, F. Picard, N. Kremer, F. Vavre, M.-F. Sagot, V. Lacroix.*SNP calling from RNA-seq data without a reference genome: identification, quantification, differential analysis and impact on the protein sequence*, in "Nucleic Acids Research", 2016 [*DOI :* 10.1093/nar/gkw655], https://hal.inria.fr/hal-01352586.

[19] T. Marschall, M. Marz, T. Abeel, L. Dijkstra, B. E. Dutilh, A. Ghaffaari, P. Kersey, W. P. Kloosterman, V. Makinen, A. M. Novak, B. Paten, D. Porubsky, E. Rivals, C. Alkan, J. A. Baaijens, P. I. W. D. Bakker, V. Boeva, R. J. P. Bonnal, F. Chiaromonte, R. Chikhi, F. D. Ciccarelli, R. Cijvat, E. Datema, C. M. V. Duijn, E. E. Eichler, C. Ernst, E. Eskin, E. Garrison, M. El-Kebir, G. W. Klau, J. O. Korbel, E.-W. Lameijer, B. Langmead, M. Martin, P. Medvedev, J. C. Mu, P. Neerincx, K. Ouwens, P. Peterlongo, N. Pisanti, S. Rahmann, B. Raphael, K. Reinert, D. d. Ridder, J. d. Ridder, M. Schlesner, O. Schulz-Trieglaff, A. D. Sanders, S. Sheikhizadeh, C. Shneider, S. Smit, D. Valenzuela, J. Wang, L. Wessels, Y. Zhang, V. Guryev, F. Vandin, K. Ye, A. Schönhuth.*Computational pan-genomics: status, promises and challenges*, in "Briefings in Bioinformatics", October 2016 [*DOI :* 10.1093/bib/bbw089], https://hal.inria.fr/hal-01390478.

[20] A. MEYGRET, V. PASCAL, S. MOULLEC, J. NACAZUME, Y. ADNANI, D. LAVENIER, S. KAYAL, A. FAILI.*Genome sequence of the uncommon Streptococcus pyogenes M/emm66 strain STAB13021, isolated from clonal clustered cases in French Brittany*, in "Genome Announcements", July 2016, vol. 4, n^{o} 4, 16, e00689 [*DOI :* 10.1128/GENOMEA.00689-16], https://hal.inria.fr/hal-01385659.

[21] C. RISPE, F. LEGEAI, D. PAPURA, A. BRETAUDEAU, S. HUDAVERDIAN, G. LE TRIONNAIRE, D. TAGU, J. JAQUIÉRY, F. DELMOTTE.*De novo transcriptome assembly of the grapevine phylloxera allows identification of genes differentially expressed between leaf- and root-feeding forms*, in "BMC Genomics", December 2016, vol. 17, n^{o} 1, 219 [*DOI :* 10.1186/S12864-016-2530-8], https://hal.inria.fr/hal-01286528.

International Conferences with Proceedings

[22] S. FIDANOVA, O. ROEVA, A. MUCHERINO, K. KAPANOVA.*InterCriteria Analysis of Ant Algorithm with Environment Change for GPS Surveying Problem*, in "17th International Conference on Artificial Intelligence: Methodology , Systems, Applications (AIMSA16)", Varna, Bulgaria, C. DICHEV, G. AGRE (editors), Lecture Notes in Artificial Intelligence, Springer, September 2016, vol. 9883, p. 271–278 [*DOI :* 10.1007/978-3-319-44748-3_26], https://hal.inria.fr/hal-01402412.

[23] W. GRAMACHO, A. MUCHERINO, J.-H. LIN, C. LAVOR.*A New Approach to the Discretization of Multidimensional Scaling*, in "IEEE Conference Proceedings of FedCSIS16", Gandz, Poland, September 2016, https://hal.inria.fr/hal-01402390.

[24] C. MARCHET, A. LIMASSET, L. BITTNER, P. PETERLONGO.*A resource-frugal probabilistic dictionary and applications in (meta)genomics*, in "Prageu Stringology Conference", Prague, Czech Republic, August 2016, https://hal.inria.fr/hal-01386744.

Conferences without Proceedings

[25] S. FRANÇOIS, R. ANDONOV, H. DJIDJEV, D. LAVENIER.*Global Optimization Methods for Genome Scaffolding*, in "12th International Workshop on Constraint-Based Methods for Bioinformatics", Toulouse, France, September 2016, https://hal.inria.fr/hal-01385665.

[26] D. LAVENIER, J.-F. ROY, D. FURODET.*DNA Mapping using Processor-in-Memory Architecture*, in "Workshop on Accelerator-Enabled Algorithms and Applications in Bioinformatics", Shenzhen, China, December 2016, https://hal.archives-ouvertes.fr/hal-01399997.

[27] F. LEGEAI, C. BETTEMBOURG, A. BRETAUDEAU, Y. CHAUSSIN, O. DAMERON, D. TAGU.*BIPAA/Askomics, a new and easy approach for querying genomics and epigenomics elements in interaction*, in "XXVth International Congress of Entomology 2016", Orlando, Florida, United States, September 2016, https://hal.inria.fr/hal-01391080.

[28] H.-S. PHAM, D. LAVENIER, A. TERMIER.*Identifying Genetic Variant Combinations using Skypatterns*, in "7th International Workshop on Biological Knowledge Discovery and Data Mining (Workshop BIOKDD '16)", Porto, Portugal, DEXA, September 2016 [*DOI :* 10.1109/DEXA.2016.13], https://hal.inria.fr/hal-01385614.

Scientific Books (or Scientific Book chapters)

[29] J. A. BRISSON, J. JAQUIÉRY, F. LEGEAI, G. L. TRIONNAIRE, D. TAGU.*Genomics of Phenotypic plastity in Aphids*, in "Management of Insect Pests to Agriculture", H. CZOSNEK, M. GHANIM (editors), Springer

International Publishing, 2016, p. 65-96 [*DOI :* 10.1007/978-3-319-24049-7_3], https://hal-univ-rennes1.archives-ouvertes.fr/hal-01314988.

[30] G. Chapuis, H. Djidjev, D. Lavenier, R. Andonov.*GPU-accelerated shortest paths computations for planar graphs*, in "Advances in GPU Research and Practice", H. Azad (editor), Elsevier, September 2016, 774, https://hal.inria.fr/hal-01385634.

[31] A. Mucherino, S. Fidanova, M. Ganzha.*Introducing the Environment in Ant Colony Optimization*, in "Studies in Computational Intelligence", Recent Advances in Computational Optimization, Springer, July 2016, vol. 655, p. 147–158, https://hal.inria.fr/hal-01402423.

Research Reports

[32] L. Bouri, D. Lavenier.*Evaluation des logiciels d'assemblage utilisant des lectures longues*, Inria Rennes - Bretagne Atlantique, November 2016, n^{o} RT-0475, https://hal.inria.fr/hal-01282892.

[33] D. Lavenier, C. Deltel, D. Furodet, J.-F. Roy.*BLAST on UPMEM*, Inria Rennes - Bretagne Atlantique, March 2016, n^{o} RR-8878, 20, https://hal.archives-ouvertes.fr/hal-01294345.

[34] D. Lavenier, C. Deltel, D. Furodet, J.-F. Roy.*MAPPING on UPMEM*, Inria, June 2016, n^{o} RR-8923, 17, https://hal.archives-ouvertes.fr/hal-01327511.

Other Publications

[35] G. Benoit, P. Peterlongo, M. Mariadassou, E. Drezen, S. Schbath, D. Lavenier, C. Lemaitre.*Multiple Comparative Metagenomics using Multiset k-mer Counting*, April 2016, working paper or preprint, https://hal.inria.fr/hal-01300485.

[36] A. Evrard, C. Bettembourg, M. M. Jubault, O. Dameron, O. Filangi, A. Bretaudeau, F. F. Legeai.*Integration and query of biological datasets with Semantic Web technologies: AskOmics*, June 2016, Journées Ouvertes Biologie, Informatique et Mathématiques (JOBIM 2016), Poster, https://hal.inria.fr/hal-01391087.

[37] A. Limasset, C. Marchet, P. Peterlongo, L. Bittner.*Minimal perfect hash functions in large scale bioinformatics Problem*, June 2016, JOBIM 2016, Poster, https://hal.archives-ouvertes.fr/hal-01341718.

[38] C. Marchet, A. Limasset, L. Bittner, P. Peterlongo.*A resource-frugal probabilistic dictionary and applications in (meta)genomics*, May 2016, working paper or preprint, https://hal.inria.fr/hal-01322440.

[39] F. Moreews, D. Lavenier.*Seamless Coarse Grained Parallelism Integration in Intensive Bioinformatics Workflows*, May 2016, working paper or preprint [*DOI :* 10.1145/2488551.2488588], https://hal.inria.fr/hal-00908842.

Project-Team HYBRID

3D interaction with virtual environments using body and mind

IN PARTNERSHIP WITH:
Institut national des sciences appliquées de Rennes

Université Rennes 1

RESEARCH CENTER
Rennes - Bretagne-Atlantique

THEME
Interaction and visualization

Table of contents

Project-Team HYBRID

Creation of the Team: 2013 January 01, updated into Project-Team: 2013 July 01

Keywords:

Computer Science and Digital Science:

2.5. - Software engineering
5. - Interaction, multimedia and robotics
5.1. - Human-Computer Interaction
5.1.2. - Evaluation of interactive systems
5.1.3. - Haptic interfaces
5.1.4. - Brain-computer interfaces, physiological computing
5.1.5. - Body-based interfaces
5.1.7. - Multimodal interfaces
5.1.8. - 3D User Interfaces
5.5.4. - Animation
5.6. - Virtual reality, augmented reality
6. - Modeling, simulation and control
6.2. - Scientific Computing, Numerical Analysis & Optimization
6.3. - Computation-data interaction

Other Research Topics and Application Domains:

1.3. - Neuroscience and cognitive science
2. - Health
2.4. - Therapies
2.5. - Handicap and personal assistances
2.6. - Biological and medical imaging
2.7. - Medical devices
2.7.1. - Surgical devices
2.8. - Sports, performance, motor skills
5. - Industry of the future
5.1. - Factory of the future
5.2. - Design and manufacturing
5.8. - Learning and training
5.9. - Industrial maintenance
8.1. - Smart building/home
8.3. - Urbanism and urban planning
9.1. - Education
9.2. - Art
9.2.2. - Cinema, Television
9.2.3. - Video games
9.3. - Sports
9.5.6. - Archeology, History

1. Members

Research Scientists

Anatole Lécuyer [Team leader, Inria, Senior Researcher, HDR]
Fernando Argelaguet Sanz [Inria, Researcher]

Faculty Members

Bruno Arnaldi [INSA Rennes, Professor, HDR]
Valérie Gouranton [INSA Rennes, Associate Professor]
Maud Marchal [INSA Rennes, Associate Professor, HDR]

Technical Staff

Ronan Gaugne [Univ. Rennes 1, SED Research Engineer 15%]
Florian Nouviale [INSA Rennes, SED Research Engineer 20%]
Jérôme Chabrol [Inria, until Oct 2016]
Guillaume Claude [INSA Rennes, from Oct 2016]
Charles Garraud [Inria, until Oct 2016]
Francois Lehericey [VINCI, from Nov 2016]
Jussi Tapio Lindgren [Inria]
Marsel Mano [Inria]

PhD Students

Jean-Baptiste Barreau [CNRS]
Guillaume Claude [INSA Rennes, until Sept 2016]
Guillaume Cortes [Realyz]
Antoine Costes [Technicolor]
Anne-Solène Dris-Kerdreux [VINCI]
Andéol Evain [Inria]
Jérémy Lacoche [IRT B-COM, until Aug 2016]
Morgan Le Chénéchal [IRT B-COM, until Sep 2016]
Benoît Le Gouis [INSA Rennes]
Gwendal Le Moulec [INSA Rennes]
Francois Lehericey [INSA Rennes, until Oct 2016]
Lorraine Perronnet [Inria]
Gautier Picard [INSA Rennes, from Oct 2016]
Hakim Si Mohammed [Inria, from Oct 2016]

Post-Doctoral Fellows

Kevin - Yoren Gaffary [Inria]
Adrien Girard [Inria, until May 2016]
Nataliya Kos'Myna [Inria]

Administrative Assistant

Nathalie Denis [Inria]

2. Overall Objectives

2.1. Overall Objectives

Our research project belongs to the scientific field of Virtual Reality (VR) and 3D interaction with virtual environments. VR systems can be used in numerous applications such as for industry (virtual prototyping, assembly or maintenance operations, data visualization), entertainment (video games, theme parks), arts and design (interactive sketching or sculpture, CAD, architectural mock-ups), education and science (physical simulations, virtual classrooms), or medicine (surgical training, rehabilitation systems). A major change that we foresee in the next decade concerning the field of Virtual Reality relates to the emergence of new paradigms of interaction (input/output) with Virtual Environments (VE).

As for today, the most common way to interact with 3D content still remains by measuring user's motor activity, i.e., his/her gestures and physical motions when manipulating different kinds of input device. However, a recent trend consists in soliciting more movements and more physical engagement of the body of the user. We can notably stress the emergence of bimanual interaction, natural walking interfaces, and whole-body involvement. These new interaction schemes bring a new level of complexity in terms of generic physical simulation of potential interactions between the virtual body and the virtual surrounding, and a challenging "trade-off" between performance and realism. Moreover, research is also needed to characterize the influence of these new sensory cues on the resulting feelings of "presence" and immersion of the user.

Besides, a novel kind of user input has recently appeared in the field of virtual reality: the user's mental activity, which can be measured by means of a "Brain-Computer Interface" (BCI). Brain-Computer Interfaces are communication systems which measure user's electrical cerebral activity and translate it, in real-time, into an exploitable command. BCIs introduce a new way of interacting "by thought" with virtual environments. However, current BCI can only extract a small amount of mental states and hence a small number of mental commands. Thus, research is still needed here to extend the capacities of BCI, and to better exploit the few available mental states in virtual environments.

Our first motivation consists thus in designing novel “body-based” and “mind-based” controls of virtual environments and reaching, in both cases, more immersive and more efficient 3D interaction.

Furthermore, in current VR systems, motor activities and mental activities are always considered separately and exclusively. This reminds the well-known “body-mind dualism” which is at the heart of historical philosophical debates. In this context, our objective is to introduce novel “hybrid” interaction schemes in virtual reality, by considering motor and mental activities jointly, i.e., in a harmonious, complementary, and optimized way. Thus, we intend to explore novel paradigms of 3D interaction mixing body and mind inputs. Moreover, our approach becomes even more challenging when considering and connecting multiple users which implies multiple bodies and multiple brains collaborating and interacting in virtual reality.

Our second motivation consists thus in introducing a “hybrid approach” which will mix mental and motor activities of one or multiple users in virtual reality.

3. Research Program

3.1. Research Program

The scientific objective of Hybrid team is to improve 3D interaction of one or multiple users with virtual environments, by making full use of physical engagement of the body, and by incorporating the mental states by means of brain-computer interfaces. We intend to improve each component of this framework individually, but we also want to improve the subsequent combinations of these components.

The "hybrid" 3D interaction loop between one or multiple users and a virtual environment is depicted in Figure 1. Different kinds of 3D interaction situations are distinguished (red arrows, bottom): 1) body-based interaction, 2) mind-based interaction, 3) hybrid and/or 4) collaborative interaction (with at least two users). In each case, three scientific challenges arise which correspond to the three successive steps of the 3D interaction loop (blue squares, top): 1) the 3D interaction technique, 2) the modeling and simulation of the 3D scenario, and 3) the design of appropriate sensory feedback.

The 3D interaction loop involves various possible inputs from the user(s) and different kinds of output (or sensory feedback) from the simulated environment. Each user can involve his/her body and mind by means of corporal and/or brain-computer interfaces. A hybrid 3D interaction technique (1) mixes mental and motor inputs and translates them into a command for the virtual environment. The real-time simulation (2) of the virtual environment is taking into account these commands to change and update the state of the virtual world and virtual objects. The state changes are sent back to the user and perceived by means of different sensory feedbacks (e.g., visual, haptic and/or auditory) (3). The sensory feedbacks are closing the 3D interaction loop. Other users can also interact with the virtual environment using the same procedure, and can eventually “collaborate” by means of “collaborative interactive techniques” (4).

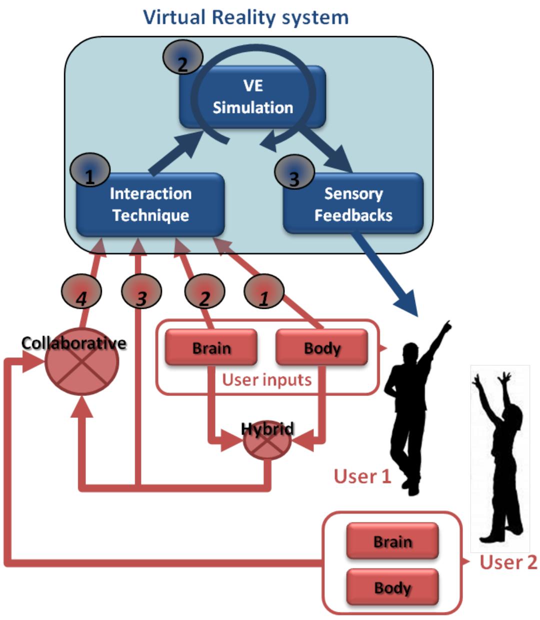


Figure 1. 3D hybrid interaction loop between one or multiple users and a virtual reality system. Top (in blue) three steps of 3D interaction with a virtual environment: (1-blue) interaction technique, (2-blue) simulation of the virtual environment, (3-blue) sensory feedbacks. Bottom (in red) different cases of interaction: (1-red) body-based, (2-red) mind-based, (3-red) hybrid, and (4-red) collaborative 3D interaction.

This description is stressing three major challenges which correspond to three mandatory steps when designing 3D interaction with virtual environments:

- **3D interaction techniques:** This first step consists in translating the actions or intentions of the user (inputs) into an explicit command for the virtual environment. In virtual reality, the classical tasks that require such kinds of user command were early categorized in four [38]: navigating the virtual world, selecting a virtual object, manipulating it, or controlling the application (entering text, activating options, etc). The addition of a third dimension, the use of stereoscopic rendering and the use of advanced VR interfaces make however inappropriate many techniques that proved efficient in 2D, and make it necessary to design specific interaction techniques and adapted tools. This challenge is here renewed by the various kinds of 3D interaction which are targeted. In our case, we consider various cases, with motor and/or cerebral inputs, and potentially multiple users.
- **Modeling and simulation of complex 3D scenarios:** This second step corresponds to the update of the state of the virtual environment, in real-time, in response to all the potential commands or actions sent by the user. The complexity of the data and phenomena involved in 3D scenarios is constantly increasing. It corresponds for instance to the multiple states of the entities present in the simulation (rigid, articulated, deformable, fluids, which can constitute both the user's virtual body and the different manipulated objects), and the multiple physical phenomena implied by natural human interactions (squeezing, breaking, melting, etc). The challenge consists here in modeling and simulating these complex 3D scenarios and meeting, at the same time, two strong constraints of virtual reality systems: performance (real-time and interactivity) and genericity (e.g., multi-resolution, multi-modal, multi-platform, etc).
- **Immersive sensory feedbacks:** This third step corresponds to the display of the multiple sensory feedbacks (output) coming from the various VR interfaces. These feedbacks enable the user to perceive the changes occurring in the virtual environment. They are closing the 3D interaction loop, making the user immersed, and potentially generating a subsequent feeling of presence. Among the various VR interfaces which have been developed so far we can stress two kinds of sensory feedback: visual feedback (3D stereoscopic images using projection-based systems such as CAVE systems or Head Mounted Displays); and haptic feedback (related to the sense of touch and to tactile or force-feedback devices). The Hybrid team has a strong expertize in haptic feedback, and in the design of haptic and "pseudo-haptic" rendering [41]. Note that a major trend in the community, which is strongly supported by the Hybrid team, relates to a "perception-based" approach, which aims at designing sensory feedbacks which are well in line with human perceptual capacities.

These three scientific challenges are addressed differently according to the context and the user inputs involved. We propose to consider three different contexts, which correspond to the three different research axes of the Hybrid research team, namely : 1) body-based interaction (motor input only), 2) mind-based interaction (cerebral input only), and then 3) hybrid and collaborative interaction (i.e., the mixing of body and brain inputs from one or multiple users).

3.2. Research Axes

The scientific activity of Hybrid team follows three main axes of research:

- **Body-based interaction in virtual reality.** Our first research axis concerns the design of immersive and effective "body-based" 3D interactions, i.e., relying on a physical engagement of the user's body. This trend is probably the most popular one in VR research at the moment. Most VR setups make use of tracking systems which measure specific positions or actions of the user in order to interact with a virtual environment. However, in recent years, novel options have emerged for measuring "full-body" movements or other, even less conventional, inputs (e.g. body equilibrium). In this first research axis we are thus concerned by the emergence of new kinds of "body-based interaction" with virtual environments. This implies the design of novel 3D user interfaces and novel 3D interactive techniques, novel simulation models and techniques, and novel sensory feedbacks for body-based interaction with virtual worlds. It involves real-time physical simulation of complex interactive phenomena, and the design of corresponding haptic and pseudo-haptic feedback.

- **Mind-based interaction in virtual reality.** Our second research axis concerns the design of immersive and effective “mind-based” 3D interactions in Virtual Reality. Mind-based interaction with virtual environments is making use of Brain-Computer Interface technology. This technology corresponds to the direct use of brain signals to send “mental commands” to an automated system such as a robot, a prosthesis, or a virtual environment. BCI is a rapidly growing area of research and several impressive prototypes are already available. However, the emergence of such a novel user input is also calling for novel and dedicated 3D user interfaces. This implies to study the extension of the mental vocabulary available for 3D interaction with VE, then the design of specific 3D interaction techniques "driven by the mind" and, last, the design of immersive sensory feedbacks that could help improving the learning of brain control in VR.
- **Hybrid and collaborative 3D interaction.** Our third research axis intends to study the combination of motor and mental inputs in VR, for one or multiple users. This concerns the design of mixed systems, with potentially collaborative scenarios involving multiple users, and thus, multiple bodies and multiple brains sharing the same VE. This research axis therefore involves two interdependent topics: 1) collaborative virtual environments, and 2) hybrid interaction. It should end up with collaborative virtual environments with multiple users, and shared systems with body and mind inputs.

4. Application Domains

4.1. Overview

The research program of Hybrid team aims at next generations of virtual reality and 3D user interfaces which could possibly address both the “body” and “mind” of the user. Novel interaction schemes are designed, for one or multiple users. We target better integrated systems and more compelling user experiences.

The applications of our research program correspond to the applications of virtual reality technologies which could benefit from the addition of novel body-based or mind-based interaction capabilities:

- **Industry**: with training systems, virtual prototyping, or scientific visualization;
- **Medicine**: with rehabilitation and reeducation systems, or surgical training simulators;
- **Entertainment**: with 3D web navigations, video games, or attractions in theme parks,
- **Construction**: with virtual mock-ups design and review, or historical/architectural visits.

5. Highlights of the Year

5.1. Highlights of the Year

- Two new permanent staff have joined our team this year: Ronan Gaugne (Research Engineer, Univ. Rennes 1), Ferran Argelaguet (CR2 Inria Research Scientist).
- There has been an outstanding total of six PhD Theses defended this year by members of Hybrid.
- Our team organized, together with MimeTIC team, a press conference in Paris on the "6-Finger Illusion" in May 2016, followed by a huge media coverage.

5.1.1. Awards

- Paper and demo "When the Giant meets the Ant: An Asymmetric Approach for Collaborative and Concurrent Object Manipulation in a Multi-Scale Environment" [35] obtained the Second Prize at the IEEE 3DUI Contest 2016.
- Project MANDARIN received the "Economical Impact Award 2016" from ANR (French National Research Agency).

- Project PREVIZ received a "Loading the Future' Trophy 2016" from Images et Réseaux French Competitivity Cluster.

6. New Software and Platforms

6.1. #FIVE

KEYWORDS: Virtual reality - Behaviour - 3D interaction
FUNCTIONAL DESCRIPTION

#FIVE (Framework for Interactive Virtual Environments) is a framework for the development of interactive and collaborative virtual environments. #FIVE was developed to answer the need for an easier and a faster design and development of virtual reality applications. #FIVE provides a toolkit that simplifies the declaration of possible actions and behaviours of objects in a VE. It also provides a toolkit that facilitates the setting and the management of collaborative interactions in a VE. It is compliant with a distribution of the VE on different setups. It also proposes guidelines to efficiently create a collaborative and interactive VE. The current implementation is in C# and comes with a Unity3D engine integration, compatible with MiddleVR framework. #FIVE contains software modules that can be interconnected and helps in building interactive and collaborative virtual environments. The user can focus on domain-specific aspects for his/her application thanks to #FIVE's modules. These modules can be used in a vast range of domains using virtual reality applications and requiring interactive environments and collaboration, such as in training for example.

- Participants: Bruno Arnaldi, Valerie Gouranton, Florian Nouviale, Guillaume Claude
- Contact: Valerie Gouranton and Florian Nouviale
- URL: https://bil.inria.fr/fr/software/view/2527/tab

6.2. #SEVEN

KEYWORDS: Virtual reality - Scenario - Training - Petri Net - 3D interaction
FUNCTIONAL DESCRIPTION

#SEVEN (Sensor Effector Based Scenarios Model for Driving Collaborative Virtual Environments) is a sensor effector based scenario engine that enables the execution of complex scenarios for driving Virtual Reality applications. #SEVEN's scenarios are based on an enhanced Petri net model which is able to describe and solve intricate event sequences. #SEVEN comes with an editor for creating, editing and remotely controlling and running scenarios. #SEVEN is implemented in C# and can be used as a stand-alone application or as a library. An integration to the Unity3D engine, compatible with MiddleVR, also exists.

- Participants: Bruno Arnaldi, Valerie Gouranton, Florian Nouviale, Guillaume Claude
- Contact: Valerie Gouranton and Florian Nouviale
- URL: https://bil.inria.fr/fr/software/view/2528/tab

6.3. OpenViBE

KEYWORDS: Neurosciences - Interaction - Virtual reality - Health - Real time - Neurofeedback - Brain-Computer Interface - EEG - 3D interaction
FUNCTIONAL DESCRIPTION

OpenViBE is a free and open-source software platform devoted to the design, test and use of Brain-Computer Interfaces (BCI). The platform consists of a set of software modules that can be integrated easily and efficiently to design BCI applications. The key features of OpenViBE software are its modularity, its high-performance, its portability, its multiple-users facilities and its connection with high-end/VR displays. The designer of the platform enables to build complete scenarios based on existing software modules using a dedicated graphical language and a simple Graphical User Interface (GUI). This software is available on the Inria Forge under the terms of the AGPL licence, and it was officially released in June 2009. Since then, the OpenViBE software has already been downloaded more than 40000 times, and it is used by numerous laboratories, projects, or individuals worldwide. More information, downloads, tutorials, videos, documentations are available on the OpenViBE website.

- Participants: Anatole Lécuyer, Jussi Tapio Lindgren, Jerome Chabrol, Charles Garraud, and Marsel Mano
- Partners: Inria teams POTIOC, ATHENA and NEUROSYS
- Contact: Anatole Lécuyer
- URL: http://openvibe.inria.fr
- URL: https://bil.inria.fr/fr/software/view/1194/tab

6.4. Platform: Immerstar

- Participants : Florian Nouviale, Ronan Gaugne

With the two platforms of virtual reality, Immersia and Immermove, grouped under the name Immerstar, the team has access to high level scientific facilities. This equipment benefits the research teams of the center and has allowed them to extend their local, national and international collaborations. The Immerstar platform is granted by a Inria CPER funding for 2015-2019 that enables important evolutions of the equipment. In 2016, the first technical evolutions have been decided, with, for Immermove, the addition of a third face to the immersive space, and the extension of the Vicon tracking system, and for Immersia, the installation of WQXGA laser projectors and of a new tracking system.

7. New Results

7.1. Virtual Reality and 3D Interaction

7.1.1. Perception in Virtual Environments

With the increasing demand in consumer VR applications, the need to understand how users perceive the virtual environment and their virtual self (avatar) is becoming more and more important. In particular, with the potential of virtual reality to alter and control avatars in different ways, the user representation in the virtual world does not always necessarily match the user body structure. Besides, the study of how the users perceive their surrounding environment (e.g. depth perception) is another active field of research in VR.

The role of interaction in virtual embodiment: Effects of the virtual hand representation

Participants: Ferran Argelaguet and Anatole Lécuyer

First, we have studied how people appropriate their virtual hand representation when interacting in virtual environments [14]. In order to answer this question, we conducted an experiment studying the sense of embodiment when interacting with three different virtual hand representations (see Figure 2), each one providing a different degree of visual realism but keeping the same control mechanism. The main experimental task was a Pick-and-Place task in which participants had to grasp a virtual cube and place it to an indicated position while avoiding an obstacle (brick, barbed wire or fire). Results show that the sense of agency is stronger for less realistic virtual hands which also provide less mismatch between the participant's actions and the animation of the virtual hand. In contrast, the sense of ownership is increased for the human virtual hand which provides a direct mapping between the degrees of freedom of the real and virtual hand.

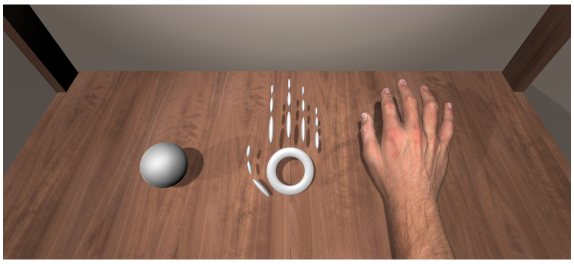


Figure 2. Evaluated virtual hand representations: abstract (left), iconic (center) and realistic virtual hands (right). Each virtual hand had its own visual feedback when the grasping operation was triggered.

This work was done in collaboration with MimeTIC team.

Wow! I Have Six Fingers!": Would You Accept Structural Changes of Your Hand in VR?

Participants: Ferran Argelaguet and Anatole Lécuyer

In a different context, we have explored how users would accept as their own a six-digit realistic virtual hand [6]. By measuring participants' senses of ownership (i.e., the impression that the virtual hand is actually our own hand) and agency (i.e., the impression to be able to control the actions of the virtual hand), we somehow evaluate the possibility of creating a Six-Finger Illusion in VR. We measured these two dimensions of virtual embodiment in a virtual reality experiment where participants performed two tasks successively: (1) a self-manipulation task inducing visuomotor feedback, where participants mimicked finger movements presented in the virtual scene and (2) a visuotactile task inspired by Rubber Hand Illusion protocols, where an experimenter stroked the hand of the user with a brush (see Figure 3). The real and virtual brushes were synchronously stroking the participants' real and virtual hand, and in the case when the virtual brush was stroking the additional virtual digit, the real ring finger was also synchronously stroked to provide consistent tactile stimulation and elicit a sense of embodiment. Results of the experiment show that participants did experience high levels of ownership and agency of the six-digit virtual hand as a whole. These results bring preliminary insights about how avatar with structural differences can affect the senses of ownership and agency experienced by users in VR.

This work was done in collaboration with MimeTIC team.

CAVE Size Matters: Effects of Screen Distance and Parallax on Distance Estimation in Large Immersive Display Setups

Participants: Ferran Argelaguet and Anatole Lécuyer

When walking within a CAVE-like system, accommodation distance, parallax, and angular resolution vary according to the distance between the user and the projection walls, which can alter spatial perception. As these systems get bigger, there is a need to assess the main factors influencing spatial perception in order to better design immersive projection systems and virtual reality applications. In this work, we performed two experiments that analyze distance perception when considering the distance toward the projection screens and parallax as main factors. Both experiments were conducted in a large immersive projection system with up to 10-meter interaction space. The first experiment showed that both the screen distance and parallax have a strong asymmetric effect on distance judgments. We observed increased underestimation for positive parallax conditions and slight distance overestimation for negative and zero parallax conditions. The second experiment further analyzed the factors contributing to these effects and confirmed the observed effects of the first experiment with a high-resolution projection setup providing twice the angular resolution and improved accommodative stimuli. In conclusion, our results suggest that space is the most important characteristic for

Figure 3. The virtual six-finger hand and the participant's hand are synchronously stimulated using a virtual and a real brush respectively.

distance perception, optimally requiring about 6- to 7-meter distance around the user, and virtual objects with high demands on accurate spatial perception should be displayed at zero or negative parallax [3].

This work was done in collaboration with MimeTIC team and the University of Hamburg.

7.1.2. 3D User Interfaces

GiAnt: stereoscopic-compliant multi-scale navigation in VEs

Participants: Ferran Argelaguet

Navigation in multi-scale virtual environments (MSVE) requires the adjustment of the navigation parameters to ensure optimal navigation experiences at each level of scale (see Figure 4). In particular, in immersive stereoscopic systems, e.g. when performing zoom-in and zoom-out operations, the navigation speed and the stereoscopic rendering parameters have to be adjusted accordingly. Although this adjustment can be done manually by the user, it can be complex, tedious and strongly depends on the virtual environment. We have proposed GiAnt (GIant/ANT) [15], a new multi-scale navigation technique which automatically and seamlessly adjusts the navigation speed and the scale factor of the virtual environment based on the user's perceived navigation speed. The adjustment ensures an almost-constant perceived navigation speed while avoiding diplopia effects or diminished depth perception due to improper stereoscopic rendering configurations. The results from the conducted user evaluation shows that GiAnt is an efficient multi-scale navigation which minimizes the changes of the scale factor of the virtual environment compared to state-of-the-art multi-scale navigation techniques.

Enjoying 360° Vision with the FlyVIZ

Participants: Florian Nouviale, Maud Marchal and Anatole Lécuyer

FlyVIZ is a novel concept of wearable display device which enables to extend the human field-of-view up to 360°. With the FlyVIZ users can enjoy an artificial omnidirectional vision and see "with eyes behind their back"! We propose a novel version of our approach called the FlyVIZ v2. It is based on affordable and on the

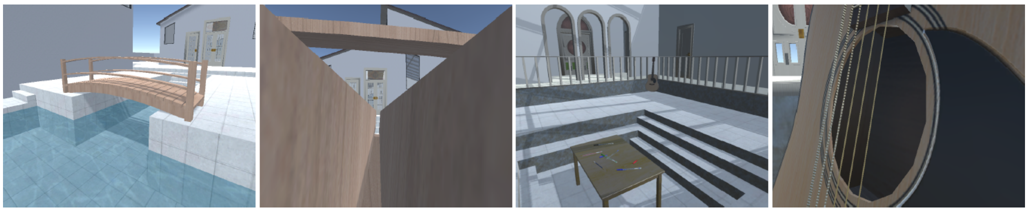


Figure 4. Multi-scale navigation sequence requiring the adaptation of the camera speed and the stereoscopic rendering parameters (e.g. parallax). GiAnt ensures that the navigation speed and the scale factor of the virtual environment are adjusted ensuring a comfortable navigation experience.

shelf components. For image acquisition, the FlyVIZ v2 relies on an iPhone4S smart-phone combined with a GoPano lens that contains a curved mirror enabling the capture of video with 360° horizontal field-of-view. For image transformation, we developed a dedicated software for iPhone that processes the video stream and transforms it into a real-time meaningful representation for the user. The "FlyVIZ_v2" was demonstrated at the ACM SIGGRAPH Emerging Technologies (2016).

Figure 5. (Left) Overview of the system. (Middle) 360° panoramic image displayed in the HMD when walking in a corridor. (Right) User grabbing an object located outside his natural field-of-view.

D3PART: A new Model for Redistribution and Plasticity of 3D User Interfaces

Participants: Jérémy Lacoche and Bruno Arnaldi

D3PART (Dynamic 3D Plastic And Redistribuable Technology) is a new model that we introduced to handle redistribution for 3D user interfaces. Redistribution consists in changing the components distribution of an interactive system across different dimensions such as platform, display and user. We extended previous plasticity models with redistribution capabilities, which lets developers create applications where 3D content and interaction tasks can be automatically redistributed across the different dimensions at runtime [21].

This work was done in collaboration with b<>com, ENIB and Telecom Bretagne.

Integration concept and model of Industry Foundation Classes (IFC) for interactive virtual environments

Participants: Anne-Solène Dris, Valérie Gouranton and Bruno Arnaldi

We defined a concept of Building Information Modeling (BIM) in combination with an integration model in order to enable interaction in Virtual Environments (see Figure 6). Such model, rich of information could be used to increase the level of abstraction of the interaction process. We proposed to explore and define how to create a BIM to ensure interoperability with the Industry Foundation Classes (IFC) model. The IFC model provides a definition of building objects, geometry, relation between objects, and other attributes such as layers, systems, link to planning, construction method, materials, domain (HVAC, Electrical, Architectural, Structure...) and quantities. The interoperability will enrich the virtual environment with the aim of creating an informed and interactive virtual environments, thus reducing the costs of applications' development. We defined a BIM modeling methodology extending the IFC interoperability to the interactive virtual environment [19].

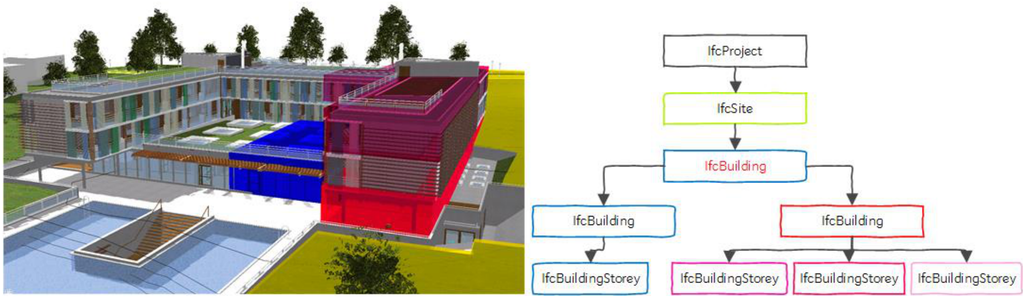


Figure 6. Interaction in virtual environments related to construction area based on BIM and a model of Industry Foundation Classes (IFC).

7.1.3. Virtual Archaeology

Digital and handcrafting processes applied to sound-studies of archaeological bone flutes

Participants: Jean-Baptiste Barreau, Ronan Gaugne, Bruno Arnaldi and Valérie Gouranton.

Bone flutes make use of a naturally hollow raw-material. As nature does not produce duplicates, each bone has its own inner cavity, and thus its own sound-potential. This morphological variation implies acoustical specificities, thus making it impossible to handcraft a true and exact sound-replica in another bone. This phenomenon has been observed in a handcrafting context and has led us to conduct two series of experiments (the first one using a handcrafting process, the second one using a 3D process) in order to investigate its exact influence on acoustics as well as on sound-interpretation based on replicas. The comparison of the results has shed light upon epistemological and methodological issues that have yet to be fully understood. This work contributes to assessing the application of digitization, 3D printing and handcrafting to flute-like sound instruments studied in the field of archaeomusicology [26].

This work was done in collaboration with MimeTIC team, ARTeHis, LBBE and Atelier El Block.

Internal 3D Printing of Intricate Structures

Participants: Ronan Gaugne, Valérie Gouranton and Bruno Arnaldi.

Additive technologies are increasingly used in Cultural Heritage process, for example in order to reproduce, complete, study or exhibit artefacts. 3D copies are based on digitization techniques such as laser scan or photogrammetry. In this case, the 3D copy remains limited to the external surface of objects. Medical images based digitization such as MRI or CT scan are also increasingly used in CH as they provide information on the internal structure of archaeological material. Different previous works illustrated the interest of combining 3D printing and CT scan in order to extract concealed artefacts from larger archaeological material. The method was based on 3D segmentation techniques within volume data obtained by CT scan to isolate nested objects. This approach was useful to perform a digital extraction, but in some case it is also interesting to observe the internal spatial organization of an intricate object in order to understand its production process. We propose a method for the representation of a complex internal structure based on a combination of CT scan and emerging 3D printing techniques mixing colored and transparent parts [25], [11]. This method was successfully applied

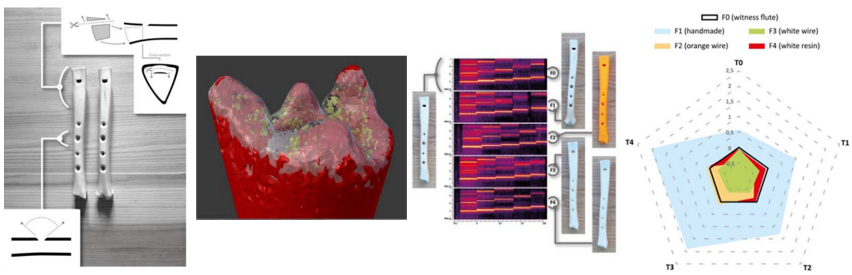


Figure 7. Sound-studies of archaeological bone flutes: (a) control flute (left) and its replica (right) both made out of goat's tibias, (b) 3D sculpted patch (transparent gray) on Blender (based on the geometry of the cloud), (c) Diagrams analysis , (d) The sound proximity of each replica comparing to the control flute, for each finger hole (numeric scale in semi-tones).

to visualize the interior of a funeral urn and is currently applied on a set of tools agglomerated in a gangue of corrosion (see Figure 8).

This work was done in collaboration with Inrap and Image ET.

Figure 8. Front and bottom views of our 3D printed urn.

7.2. Physically-Based Simulation and Multisensory Feedback

7.2.1. Physically-based Simulation

Real-time tracking of deformable targets in 3D ultrasound sequences

Participants: Maud Marchal

Soft-tissue motion tracking is an active research area that consists in providing accurate evaluation about the location of anatomical structures. To do so, ultrasound imaging is often used since it is non-invasive, real-time and portable. Thus, several ultrasound tracking approaches have been developed in order to estimate soft tissue displacements that are caused by physiological motions and manipulations by medical tools. These methods have gained significant interest for image-guided therapies such as radio-frequency ablation or high-intensity focused ultrasound. In our work, we present a real-time approach that allows tracking deformable structures in 3D ultrasound sequences [8]. Our method consists in obtaining the target displacements by combining robust dense motion estimation and mechanical model simulation. We performed an evaluation of our method through simulated data, phantom data, and real-data. Results demonstrate that this novel approach has the advantage of providing correct motion estimation regarding different ultrasound shortcomings including speckle noise, large shadows and ultrasound gain variation. Furthermore, we show the good performance of our method with respect to state-of-the-art techniques by testing on the 3D databases provided by MICCAI CLUST'14 and CLUST'15 challenges.

This work was done in collaboration with LAGADIC team and b<>com.

7.2.2. 3D Haptic Interaction

DesktopGlove: a Multi-finger Force Feedback Interface Separating Degrees of Freedom Between Hands

Participants: Merwan Achibet and Maud Marchal

In virtual environments, interacting directly with our hands and fingers greatly contributes to immersion, especially when force feedback is provided for simulating the touch of virtual objects. Yet, common haptic interfaces are unfit for multi-finger manipulation and only costly and cumbersome grounded exoskeletons do provide all the efforts expected from object manipulation. To make multi-finger haptic interaction more accessible, we have proposed to combine two affordable haptic interfaces into a bimanual setup named DesktopGlove. With this approach, each hand is in charge of different components of object manipulation: one commands the global motion of a virtual hand while the other controls its fingers for grasping (see Figure 9). In addition, each hand is subjected to forces that relate to its own degrees of freedom so that users perceive a variety of haptic effects through both of them. Our results show that (1) users are able to integrate the separated degrees of freedom of DesktopGlove to efficiently control a virtual hand in a posing task, (2) DesktopGlove shows overall better performance than a traditional data glove and is preferred by users, and (3) users considered the separated haptic feedback realistic and accurate for manipulating objects in virtual environments [12].

This work was done in collaboration with MJOLNIR team.

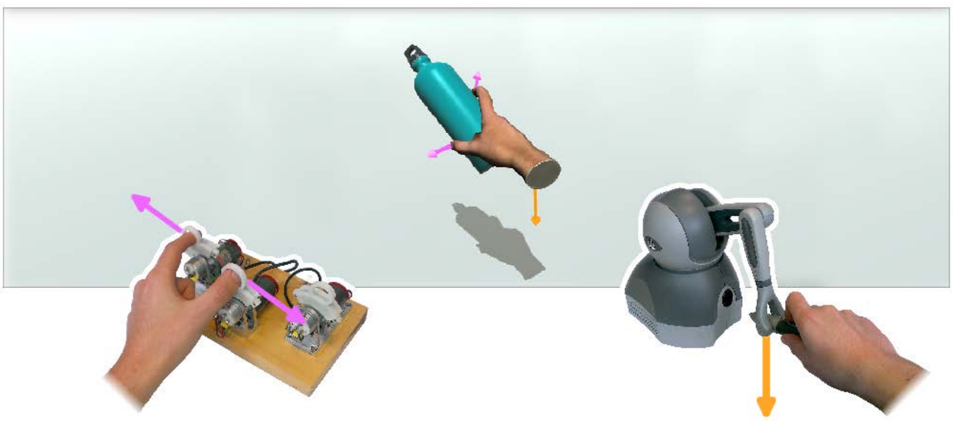


Figure 9. DesktopGlove separates the control of one virtual hand between both user's hands: a common haptic arm handles the global motion and a custom multi-finger interface controls the virtual fingers. The force feedback is split between both interfaces so that each hand is exposed to forces that relate to its own frame of reference.

ElasticArm: leveraging passive haptic feedback in virtual environments

Participants: Merwan Achibet, Adrien Girard, Anatole Lécuyer and Maud Marchal

Haptic feedback is known to improve 3D interaction in virtual environments but current haptic interfaces remain complex and tailored to desktop interaction. In [2], we describe an alternative approach called "Elastic-Arm" for incorporating haptic feedback in immersive virtual environments in a simple and cost-effective way. The Elastic-Arm is based on a body-mounted elastic armature that links the user's hand to his body and generates a progressive egocentric force when extending the arm. A variety of designs can be proposed with multiple links attached to various locations on the body in order to simulate different haptic properties and sensations such as different levels of stiffness, weight lifting, bimanual interaction, etc. Our passive haptic approach can be combined with various 3D interaction techniques and we illustrate the possibilities offered by the Elastic-Arm through several use cases based on well-known techniques such as the Bubble technique, redirected touching, and pseudo-haptics. A user study was conducted which showed the effectiveness of our pseudo-haptic technique as well as the general appreciation of the Elastic-Arm. We believe that the Elastic-Arm could be used in various VR applications which call for mobile haptic feedback or human-scale haptic sensations.

Vision-based adaptive assistance and haptic guidance for safe wheelchair corridor following

Participants: Maud Marchal

In case of motor impairments, steering a wheelchair can become a hazardous task. Joystick jerks induced by uncontrolled motions may lead to wall collisions when a user steers a wheelchair along a corridor. In [7] we introduce a low-cost assistive and guidance system for indoor corridor navigation in a wheelchair, which uses purely visual information, and which is capable of providing automatic trajectory correction and haptic guidance in order to avoid wall collisions. A visual servoing approach to autonomous corridor following serves as the backbone to this system. The algorithm employs natural image features which can be robustly extracted in real time. This algorithm is then fused with manual joystick input from the user so that progressive assistance and trajectory correction can be activated as soon as the user is in danger of collision. A force feedback in conjunction with the assistance is provided on the joystick in order to guide the user out of his dangerous trajectory. This ensures intuitive guidance and minimal interference from the trajectory correction system. In addition to being a low-cost approach, it can be seen that the proposed solution does not require an a-priori environment model. Experiments on a robotised wheelchair equipped with a monocular camera prove the capability of the system to adaptively guide and assist a user navigating in a corridor.

This work was done in collaboration with LAGADIC team.

7.2.3. Tactile Interaction at Fingertips

The fingertips are one of the most important and sensitive parts of our body. They are the first stimulated areas of the hand when we interact with our environment. Providing haptic feedback to the fingertips in virtual reality could, thus, drastically improve perception and interaction with virtual environments. Within this context, we proposed two contributions for tactile feedback and haptic interaction at the fingertips.

The Haptip

Participants: Adrien Girard, Yoren Gaffary, Anatole Lécuyer and Maud Marchal

In [5], we present a modular approach called HapTip to display such haptic sensations at the level of the fingertips. This approach relies on a wearable and compact haptic device able to simulate 2 Degree of Freedom (DoF) shear forces on the fingertip with a displacement range of 2 mm. Several modules can be added and used jointly in order to address multi-finger and/or bimanual scenarios in virtual environments. For that purpose, we introduce several haptic rendering techniques to cover different cases of 3D interaction, such as touching a rough virtual surface, or feeling the inertia or weight of a virtual object. In order to illustrate the possibilities offered by HapTip, we provide four use cases focused on touching or grasping virtual objects (see Figure 10). To validate the efficiency of our approach, we also conducted experiments to assess the tactile perception obtained with HapTip. Our results show that participants can successfully discriminate the directions of the 2 DoF stimulation of our haptic device. We found also that participants could well perceive different weights of virtual objects simulated using two HapTip devices. We believe that HapTip could be used in numerous applications in virtual reality for which 3D manipulation and tactile sensations are often crucial, such as in virtual prototyping or virtual training.

Figure 10. Illustrative use cases of our approach HapTip: the user can get in contact and tap a virtual bottle, touch a surface and feel its texture, and heft an object and feel its weight.

This work was done in collaboration with CEA List.

Studying one and two-finger perception of tactile directional cues

Participants: Yoren Gaffary, Anatole Lécuyer and Maud Marchal

In [20], we study the perception of tactile directional cues by one or two fingers, using either the index, middle, or ring finger, or any of their combination. Therefore, we use tactile devices able to stretch the skin of the fingertips in 2 DOF along four directions: horizontal, vertical, and the two diagonals. We measure the recognition rate in each direction, as well as the subjective preference, depending on the (couple of) finger(s) stimulated (see Figure 11). Our results show first that using the index and/or middle finger performs significantly better than using the ring finger on both qualitative and quantitative measures. The results when comparing one versus two-finger configurations are more contrasted. The recognition rate of the diagonals is higher when using one finger than two, whereas two fingers enable a better perception of the horizontal direction. These results pave the way to other studies on one versus two-finger perception, and raise methodological considerations for the design of multi-finger tactile devices.

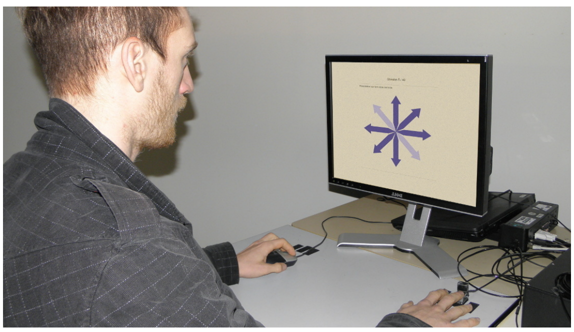


Figure 11. Experimental setup of our study on the perception of tactile directional cues by one or two fingers: the participant reports the direction of the stimulus he just perceived on the fingertip of his middle finger.

This work was done in collaboration with CEA List, IRMAR and Agrocampus Ouest.

7.3. Collaborative Virtual Environments

7.3.1. Acting in Collaborative Virtual Environments

VR Rehearsals for Acting with Visual Effects

Participants: Rozenn Bouville, Valérie Gouranton and Bruno Arnaldi,

We studied the use of Virtual Reality for movie actors rehearsals of VFX-enhanced scenes. The impediment behind VFX scenes is that actors must be filmed in front of monochromatic green or blue screens with hardly any cue to the digital scenery that is supposed to surround them. The problem is worsens when the scene includes interaction with digital partners. The actors must pretend they are sharing the set with imaginary creatures when they are, in fact, on their own on an empty set. To support actors in this complicated task, we introduced the use of VR for acting rehearsals not only to immerse actors in the digital scenery but to provide them with advanced features for rehearsing their play. Indeed, our approach combines a fully interactive environment with a dynamic scenario feature to allow actors to become familiar with the virtual elements while rehearsing dialogue and action at their own speed. The interactive and creative rehearsals enabled by the system can be either single-user or multiuser. Moreover, thanks to the wide range of supported platforms, VR rehearsals can take place either onset or offset. We conducted a preliminary study to assess whether VR training can replace classical training (see Figure 12). The results show that VR-trained actors deliver a performance just as good as ordinarily trained actors. Moreover, all the subjects in our experiment preferred VR training to classic training [17].

Figure 12. The use of VR for acting rehearsal enables actors to rehearse being immersed in the virtual scenery before being shot on a green and empty set.

Synthesis and Simulation of Collaborative Surgical Process Models

Participants: Guillaume Claude, Valérie Gouranton and Bruno Arnaldi

The use of Virtual Reality for surgical training has been mostly focused on technical surgical skills. We proposed a novel approach by focusing on the procedural aspects [4]. Our system relies on a specific work-flow, which enbables to generate a model of the procedure based on real case surgery observations made in the operating room (see Figure 13). In addition, in the context of the project S3PM we then proposed an innovative workflow to integrate the generic model of the procedure (generated from the real-case surgery observation) as a scenario model in the VR training system (see Figure 14). We described how the generic procedure model could be generated, as well as its integration in the virtual environment [18].

This work was done in collaboration with HYCOMES team and LTSI Inserm Medicis.

7.3.2. Awareness for Collaboration in Virtual Environments

Take-Over Control Paradigms in Collaborative Virtual Environments for Training

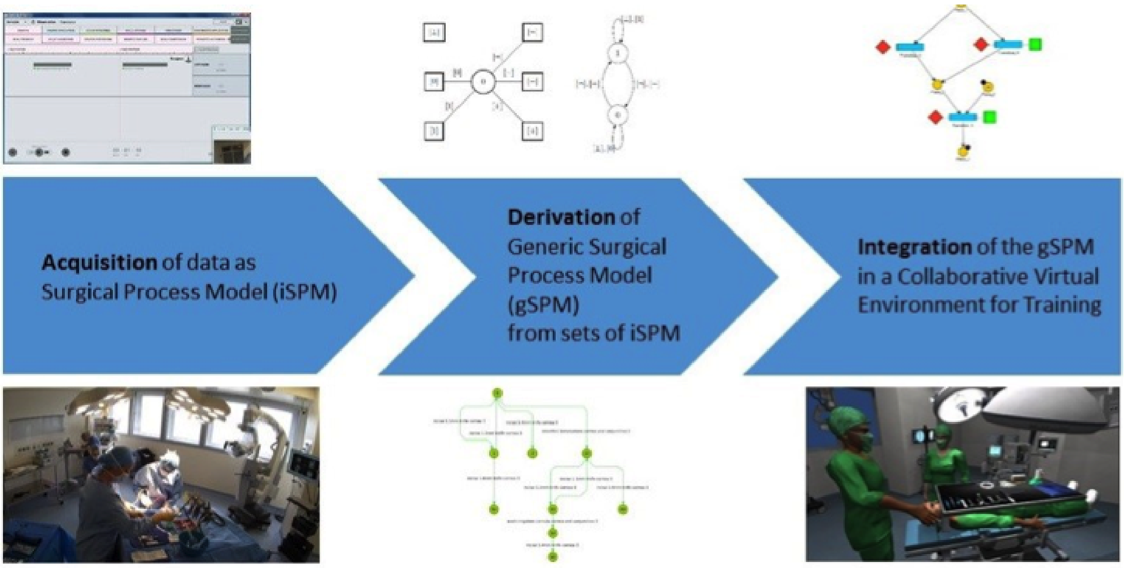
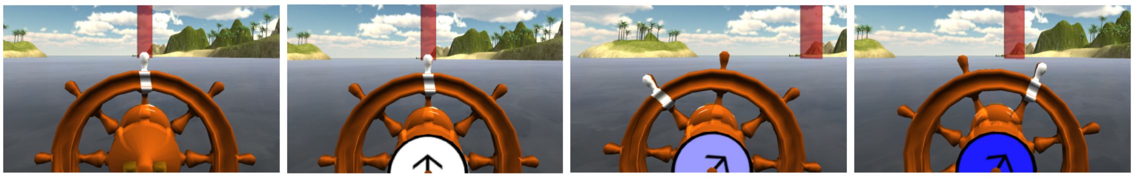


Figure 13. Collaborative Virtual Environments for Training in Surgical Procedures, based on observations during real surgeries. Observation data is integrated into a system providing a Generalised Surgical Process Model (gSPM) of the procedure. This Model is integrated as the scenario of the Virtual Environment.

Figure 14. Virtual replica of a real operating room of Rennes hospital (CHU Rennes) in the Immersia CAVE-like setup (IRISA/Inria Rennes).

Participants: Gwendal Le Moulec, Ferran Argelaguet, Anatole Lécuyer and Valérie Gouranton

We studied the notion of Take-Over Control in Collaborative Virtual Environments for Training (CVET). The Take-Over Control represents the transfer (the take over) of the interaction control of an object between two or more users. This paradigm is particularly useful for training scenarios, in which the interaction control could be continuously exchanged between the trainee and the trainer, e.g. the latter guiding and correcting the trainee's actions. We proposed a formalization of the Take-Over Control followed by an illustration focusing in a use-case of collaborative maritime navigation. In the presented use-case, the trainee has to avoid an under-water obstacle with the help of a trainer who has additional information about the obstacle. The use-case allows to highlight the different elements a Take-Over Control situation should enforce, such as user's awareness. Different Take-Over Control techniques were provided and evaluated focusing on the transfer exchange mechanism and the visual feedback (see Figure 15). The results show that participants preferred the Take-Over Control technique which maximized the user awareness [24].



(a) *Initial state* *No feedback condition* | **(b)** *Initial state* $w_1 = 1,\ w_2 = 0$ | **(c)** *In progress* $w_1 = 0.5,\ w_2 = 0.5$ | **(d)** *Completed* $w_1 = 0,\ w_2 = 1$

Figure 15. Our illustrative use case inspired by maritime navigation for Take-Over Control during training in a collaborative virtual environment. The user was instructed to steer a boat towards a semi-transparent red column (target destination) by controlling the heading of the boat. A white handle indicated the rotation angle of the boat(a). The sequence (b,c,d) shows the evolution of the contribution of the trainer on the steering angle, from no control to full control.

Vishnu: Virtual Immersive Support for HelpiNg Users: An Interaction Paradigm for Collaborative Remote Guiding in Mixed Reality

Participants: Morgan Le Chénéchal, Valérie Gouranton and Bruno Arnaldi

Increasing networking performances as well as the emergence of Mixed Reality (MR) technologies make possible providing advanced interfaces to improve remote collaboration. We presented a novel interaction paradigm called Vishnu that aims to ease collaborative remote guiding. We focus on collaborative remote maintenance as an illustrative use case. It relies on an expert immersed in Virtual Reality (VR) in the remote workspace of a local agent helped through an Augmented Reality (AR) interface. The main idea of the Vishnu paradigm is to provide the local agent with two additional virtual arms controlled by the remote expert who can use them as interactive guidance tools. Many challenges come with this: collocation, inverse kinematics (IK), the perception of the remote collaborator and gestures coordination. Vishnu aims to enhance the maintenance procedure thanks to a remote expert who can show to the local agent the exact gestures and actions to perform (see Figure 16). Our pilot user study shows that it may decrease the cognitive load compared to a usual approach based on the mapping of 2D and de-localized informations, and it could be used by agents in order to perform specific procedures without needing to have an available local expert [22].

This work was done in collaboration with b<>com and Telecom Bretagne.

When the Giant meets the Ant: An Asymmetric Approach for Collaborative and Concurrent Object Manipulation in a Multi-Scale Environment

Participants: Morgan Le Chénéchal, Jérémy Lacoche, Valérie Gouranton and Bruno Arnaldi

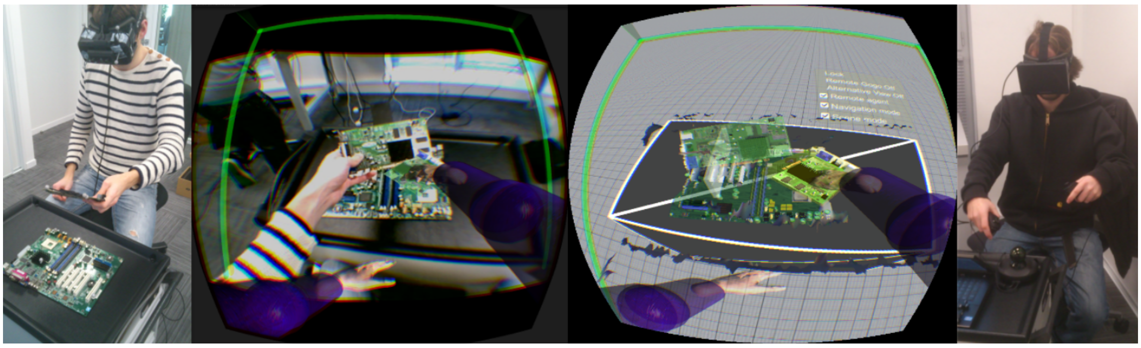


Figure 16. Illustration of the Vishnu approach: system and viewpoints of the agent (left) and the expert (right) in a motherboard assembly scenario.

We proposed a novel approach to enable two or more users to manipulate an object collaboratively. Our goal is to benefits from the wide variety of todays VR devices. Our solution is based on an asymmetric collaboration pattern at different scales in which users benefit from suited points of views and interaction techniques according to their device setups. Each user application is adapted thanks to plasticity mechanisms. Our system provides an efficient way to co-manipulate an object within irregular and narrow courses, taking advantages of asymmetric roles in synchronous collaboration (see Figure 17). Moreover, it aims to provide a way to maximize the filling of the courses while the object moves on its path [23],[35].

This work was done in collaboration with b<>com and Telecom Bretagne.

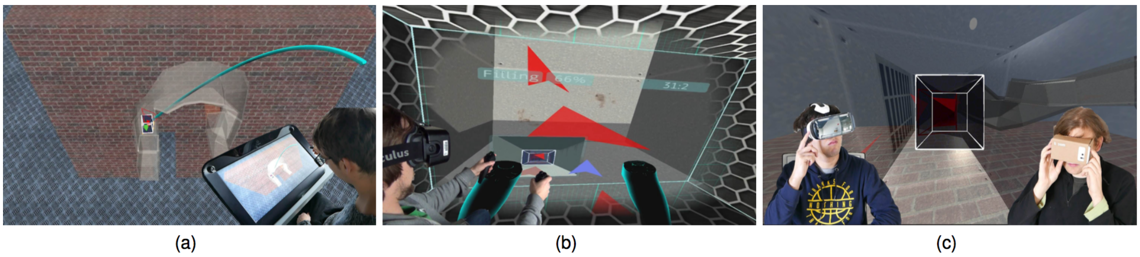


Figure 17. When the Giant meets the Ant: Collaborative manipulation of a virtual object (here, a cube) based on an asymmetric setting between two users who can be helped by two additional users. (a) The first participant has a global view of the scene and moves the object with a 3D bent ray. (b) The second user is placed inside the object and precisely rotates and scales it. (c) Two additional roles can be added. The first one helps to scale the object using a third person view of it. The other one is a spectator who switches between the other participants' viewpoints and helps them with oral communication.

7.4. Brain-Computer Interfaces

7.4.1. Contribution to a Reference Book on BCI

We have largely contributed to a reference book on BCI released in 2016 in French and English, and co-edited by Fabien Lotte, Maureen Clerc and Laurent Bougrain for ISTE (French version [36] [37]) and Wiley

(English version [39] [40]) publishers. This book provides keys for understanding and designing these multi-disciplinary interfaces, which require many fields of expertise such as neuroscience, statistics, informatics and psychology. This work corresponds to four different book chapters, all published in both French and English, which are presented hereafter.

Book chapter on BCI and videogames

Participants: Anatole Lécuyer

Videos games are often cited as a very promising field of applications for brain-computer interfaces. In a first chapter [30] [31], we described state of the art in the field of video games played "with the mind". In particular, we considered the results of the OpenViBE2 project: one of the most important research projects in this area. We presented a selection of prototypes developed during this OpenViBE2 project which is illustrative of the state of the art in this field and of the use of BCIs in video games, such as based on imagining a motion of the left and right hands to score goals, or in another example using the P300 cerebral potential to destroy spaceships in a remake of well-known Japanese game.

Book chapter on BCI softwares

Participants: Jussi Lindgren and Anatole Lécuyer

In a second chapter [28] [29], we described OpenViBE and other software platforms used to study the subject. The chapter gave an overview of such platforms. We described how the software components of the platforms reflect typical signal acquisition and signal processing stages used in BCI. Finally, we presented a high-level account of differences between major BCI platforms and gave a few pieces of advice and recommendation regarding BCI platform selection.

Book chapter on BCI and HCI

Participants: Andéol Evain, Ferran Argelaguet and Anatole Lécuyer

In a third chapter [34], we focused on the link between BCI and Human-Computer Interaction (HCI), and studied how HCI concepts can apply to BCIs. First, we presented an overview of the main concepts of HCI. We then studied the main characteristics of BCIs related to these concepts. This chapter also discussed the choice of cerebral patterns to use, depending on the interaction task and the use context. Finally, we presented the most promising new interaction paradigms for interaction with BCIs.

This work was done in collaboration with MJOLNIR team.

Book chapter on Neurofeedback

Participants: Lorraine Perronnet and Anatole Lécuyer

We proposed a fourth chapter called Brain training with Neurofeedback [33] [32]. We first defined the concept of neurofeedback (NF) and gave an overall view of the current status in this domain. Then we described the design of a NF training program and the typical course of a NF session, as well as the learning mechanisms underlying NF. We retraced the history of NF, explaining the origin of its questionable reputation and providing a foothold for understanding the diversity of existing approaches. We also discussed how the fields of NF and BCIs might potentially overlap in future with the development of "restorative" BCIs. Finally, we presented a few applications of NF and summarized the state of research of some of its major clinical applications.

This work was done in collaboration with VISAGES team.

7.4.2. BCI Methods and Techniques

Do the Stimuli of a BCI Have to be the Same as the Ones Used for Training it?

Participants: Andéol Evain, Ferran Argelaguet and Anatole Lécuyer

Does the stimulation used during the training on an SSVEP-based BCI have to be similar to that of the end use? We conducted an experiment in which we recorded six-channel EEG data from 12 subjects in various conditions of distance between targets , and of difference in color between targets [10]. Our analysis revealed that the stimulation configuration used for training which leads to the best classification accuracy is not always the one which is closest to the end use configuration. We found that the distance between targets during training is of little influence if the end use targets are close to each other, but that training at far distance can lead to a better accuracy for far distance end use. Additionally, an interaction effect is observed between training and testing color: while training with monochrome targets leads to good performance only when the test context involves monochrome targets as well, a classifier trained on colored targets can be efficient for both colored and monochrome targets. In a nutshell, in the context of SSVEP-based BCI, training using distant targets of different colors seems to lead to the best and more robust performance in all end use contexts.

This work was done in collaboration with MJOLNIR team.

A Novel Fusion Approach Combining Brain and Gaze Inputs for Target Selection

Participants: Andéol Evain, Ferran Argelaguet and Anatole Lécuyer

Gaze-based interfaces and Brain-Computer Interfaces (BCIs) allow for hands-free human-computer interaction. We investigated the combination of gaze and BCIs. We proposed a novel selection technique for 2D target acquisition based on input fusion [9]. This new approach combines the probabilistic models for each input, in order to better estimate the intent of the user. We evaluated its performance against the existing gaze and brain-computer interaction techniques. Twelve participants took part in our study, in which they had to search and select 2D targets with each of the evaluated techniques (see Figure 18). Our fusion-based hybrid interaction technique was found to be more reliable than the previous gaze and BCI hybrid interaction techniques for 10 participants over 12, while being 29% faster on average. However, similarly to what has been observed in hybrid gaze-and-speech interaction, gaze-only interaction technique still provides the best performance. Our results should encourage the use of input fusion, as opposed to sequential interaction, in order to design better hybrid interfaces.

This work was done in collaboration with MJOLNIR team.

7.4.3. BCI User Experience and Neurofeedback

Influence of Error Rate on Frustration of BCI Users

Participants: Andéol Evain, Ferran Argelaguet and Anatole Lécuyer

Brain-Computer Interfaces (BCIs) are still much less reliable than other input devices. The error rates of BCIs range from 5% up to 60%. We assessed the subjective frustration, motivation, and fatigue of BCI users, when confronted to different levels of error rate [27]. We conducted a BCI experiment in which the error rate was artificially controlled (see Figure 19). Our results first show that a prolonged use of BCI significantly increases the perceived fatigue, and induces a drop in motivation. We also found that user frustration increases with the error rate of the system but this increase does not seem critical for small differences of error rate. Thus, for future BCIs, we advise to favor user comfort over accuracy when the potential gain of accuracy remains small.

This work was done in collaboration with MJOLNIR team.

Design of an Experimental Platform for Hybrid EEG-fMRI Neurofeedback Studies

Participants: Marsel Mano, Lorraine Perronnet and Anatole Lécuyer

During a neurofeedback (NF) experiment one or more brain activity measuring technologies are used to estimate the changes of the acquired neural signals that reflect the changes of the subject's brain activity in real-time. There exist a variety of NF research applications that use only one type of neural signals (i.e. uni-modal) like EEG or fMRI, but there are very few NF researches that use two or more neural signals (i.e. multi-modal). We have developed a hybrid EEG-fMRI platform for bi-modal NF experiments, as part of the project Hemisfer. Our system is based on the integration and the synchronization of an MR-compatible EEG and fMRI acquisition subsystems. The EEG signals are acquired with a 64 channel MR-compatible solution from Brain Products and the MR imaging is performed on a 3T Verio Siemens scanner (VB17) with a 12-ch head coil. We have developed two real-time pipelines for EEG and fMRI that handle all the necessary signal processing, the Joint NF module that calculates and fuses the NF and a visualize module that displays the NF to the subject. The control and the synchronization of both subsystems with each-other and with the experimental protocol is handled by the NF Control. Our platform showed very good real-time performance

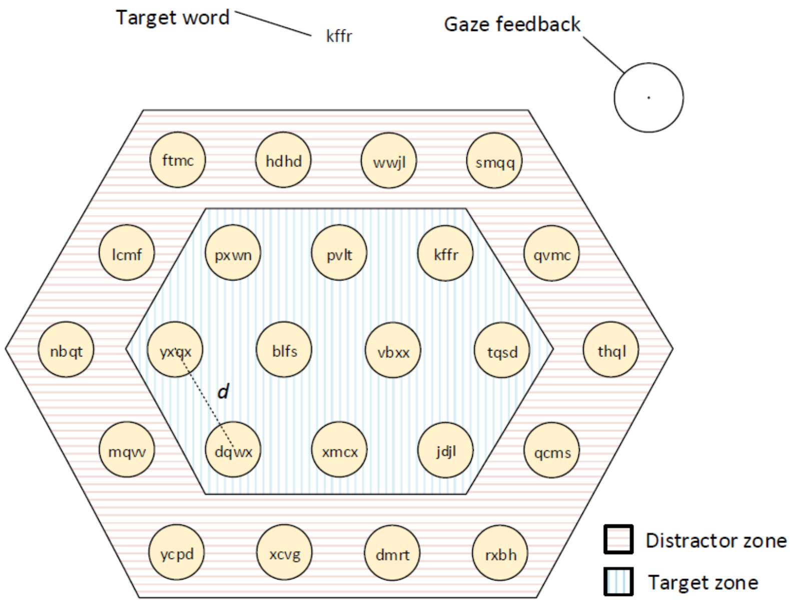


Figure 18. Experimental task combining gaze and brain inputs. The user has to look for the goal word displayed at the top of the screen, then, the user has to select the target with the exact same word. The detected gaze position is displayed under the form of a circle and a central point (visual feedback). For all trials the size of the targets remained constant, and only the length of the target word and the separation (d) between targets varied. The targets at the outer circle were distractors in which the target word was never placed.

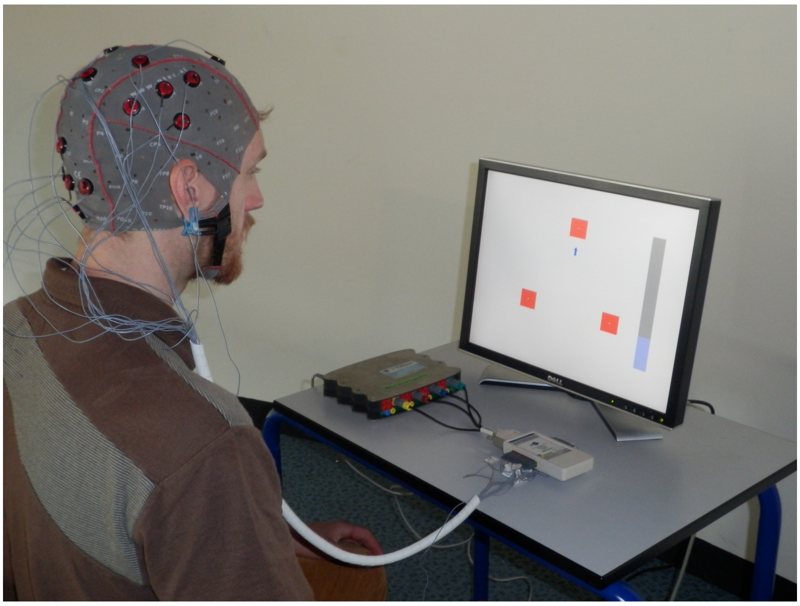


Figure 19. Experimental setup: a participant is using an SSVEP-based BCI which error rate is artificially controlled.

with various pre-processing, filtering, and NF estimation and visualization methods. The entire fMRI process from acquisition to NF takes always less than 200ms, well below the TR of regular EPI sequences (2s). The same process for EEG, with NF update cycles varying 2-5Hz, is done in virtually real time (50Hz).

This work was done in collaboration with VISAGES team and presented as poster at OHBM 2016.

Unimodal versus Bimodal EEG-fMRI Neurofeedback

Participants: Lorraine Perronnet, Anatole Lécuyer and Marsel Mano

In the context of the HEMISFER project, we proposed a simultaneous EEG-fMRI experimental protocol in which 10 healthy participants performed a motor-imagery task in unimodal and bimodal neurofeedback conditions. With this protocol we were able to compare for the first time the effects of unimodal EEG-neurofeedback and fMRI-neurofeedback versus bimodal EEG-fMRI-neurofeedback by looking both at EEG and fMRI activations. We also introduced a new feedback metaphor for bimodal EEG-fMRI-neurofeedback that integrates both EEG and fMRI signal in a single bi-dimensional feedback (a ball moving in 2D). Such a feedback is intended to relieve the cognitive load of the subject by presenting the bimodal neurofeedback task as a single regulation task instead of two. Additionally, this integrated feedback metaphor gives flexibility on defining a bimodal neurofeedback target. Participants were able to regulate activity in their motor regions in all neurofeedback conditions. Moreover, motor activations as revealed by offline fMRI analysis were stronger during EEG-fMRI-neurofeedback than during EEG-neurofeedback. This result suggests that EEG-fMRI-neurofeedback could be more specific or more engaging than EEG-neurofeedback. Our results also suggest that during EEG-fMRI-neurofeedback, participants tended to regulate more the modality that was harder to control. Taken together our results shed light on the specific mechanisms of bimodal EEG-fMRI-neurofeedback and on its added-value as compared to unimodal EEG-neurofeedback and fMRI-neurofeedback.

This work was done in collaboration with VISAGES team and presented as poster at OHBM 2016. Experiments were conducted at NEURINFO platform from University of Rennes 1.

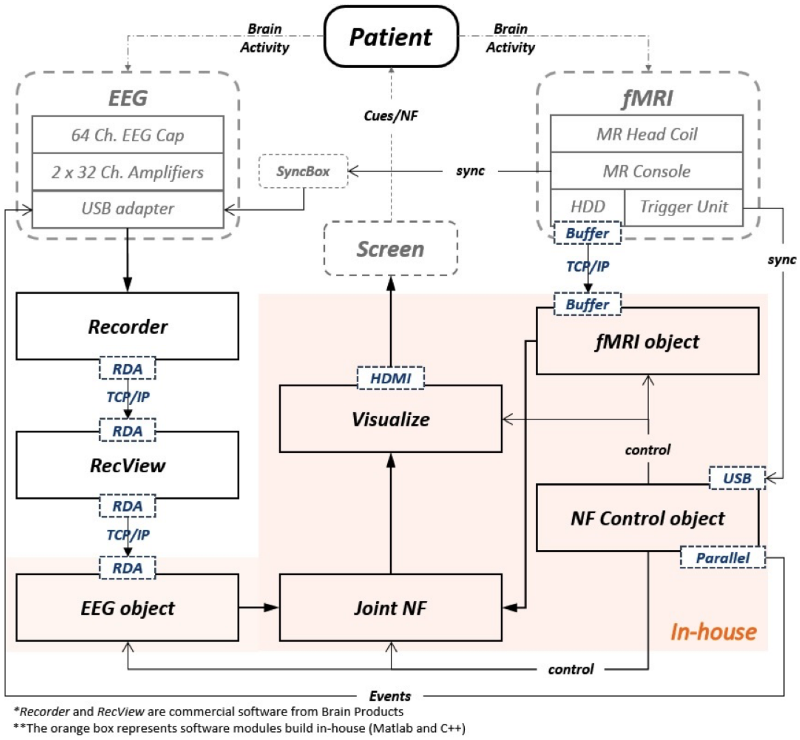


Figure 20. Architecture of our hybrid EEG-fMRI neurofeedback platform.

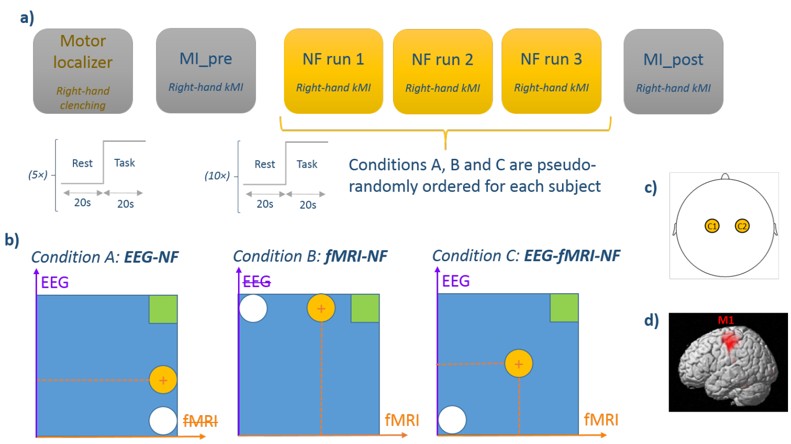


Figure 21. Experimental procedure for comparing unimodal versus bimodal EEG-fMRI neurofeedback.

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Contracts with Industry

8.1.1. *Mensia Technologies*

Participants: Anatole Lécuyer, Jussi Tapio Lindgren.

Mensia Technologies is an Inria start-up company created in November 2012 as a spin-off of Hybrid team. Mensia is focused on wellness and healthcare applications emerging from the BCI and Neurofeedback technologies. The Mensia startup should benefit from the team's expertise and of valuable and proprietary BCI research results. Mensia is based in Rennes and Paris. Anatole Lécuyer and Yann Renard (former Inria expert engineer who designed the OpenViBE software architecture and was involved in team projects for 5 years) are co-founders of Mensia Technologies together with CEO Jean-Yves Quentel.

The on-going contract between Hybrid and Mensia started in November 2013 and supports the transfer of several softwares designed by Hybrid team ("OpenViBE", "StateFinder") related to our BCI activity to Mensia Technologies for multimedia or medical applications of Mensia.

8.2. Bilateral Grants with Industry

8.2.1. *Technicolor*

Participants: Antoine Costes, Anatole Lécuyer, Ferran Argelaguet.

This grant started in December 2015. It supports Antoine Costes's CIFRE PhD program with Technicolor company on "Haptic Texturing".

8.2.2. *Realyz*

Participants: Guillaume Cortes, Anatole Lécuyer.

This grant started in December 2015. It supports Guillaume Cortes's CIFRE PhD program with Realyz company on "Improving tracking in VR".

8.2.3. *VINCI Construction*

Participants: Anne-Solène Dris-Kerdreux, Bruno Arnaldi, Valérie Gouranton.

This grant started in November 2015. It supports Anne-Solene Dris-Kerdreux's CIFRE PhD program with Vinci company on "Training in VR for construction applications".

9. Partnerships and Cooperations

9.1. Regional Initiatives

9.1.1. *Labex Cominlabs SUNSET*

Participants: Bruno Arnaldi, Guillaume Claude, Gautier Picard, Valérie Gouranton [contact].

SUNSET is a 4-year Labex Cominlabs project (2016-2020). SUNSET partners are MediCIS-LTSI (coordinator), Hybrid, Hycomes (IRISA/Inria), and CHU Rennes. SUNSET aims at developing an innovative training software suite based on immersive and collaborative virtual reality technology for training and evaluating non-technical skills. This approach will be implemented and evaluated in the context of training neurosurgical scrub nurses. We will notably integrate methods and systems developed in the S3PM project (see bellow). By relying on Human Factors approaches, the project also addresses training and evaluation of interpersonal skills. Whereas the developed technologies and approaches will be generic and adaptable to any surgical specialty, the project will evaluate the developed system within training sessions performed with scrub nurses. We ambition to propose novel approaches for surgical non-technical skill learning and assessment, and to install the developed training factory at the University Hospital of Rennes, and evaluate it with real-scale user studies.

9.1.2. Labex Cominlabs S3PM

Participants: Bruno Arnaldi, Guillaume Claude, Valérie Gouranton [contact].

S3PM ("Synthesis and Simulation of Surgical Process Models") is a 4-year Labex Cominlabs project (2013-2017). S3PM partners are MediCIS-LTSI (coordinator), Hybrid, Hycomes (IRISA/Inria), and CHU Rennes. The objective of S3PM is to propose a solution for the computation of surgical procedural knowledge models from recordings of individual procedures, and their execution. The goal of the Hybrid team is to propose and use new models for collaborative and interactive virtual environments for procedural training. The Hybrid team also works on the creation of a surgical training application in virtual reality, exposing the different contributions. Ar

9.1.3. Labex Cominlabs HEMISFER

Participants: Anatole Lécuyer [contact], Marsel Mano, Lorraine Perronnet.

HEMISFER is a 4-year project (2013-2017) funded by Labex CominLabs. It involves 4 Inria/IRISA teams (Hybrid, Visages (lead), Panama, Athena) and 2 medical centers: the Rennes Psychiatric Hospital (CHGR) and the Reeducation Department of Rennes Hospital (CHU Pontchaillou). The goal of HEMISFER is to make full use of neurofeedback paradigm in the context of rehabilitation and psychiatric disorders. The major breakthrough will come from the use of a coupling model associating functional and metabolic information from Magnetic Resonance Imaging (fMRI) to Electro-encephalography (EEG) to "enhance" the neurofeedback protocol. Clinical applications concern motor, neurological and psychiatric disorders (stroke, attention-deficit disorder, treatment-resistant mood disorders, etc).

9.1.4. Labex Cominlabs SABRE

Participants: Anatole Lécuyer [contact], Jussi Tapio Lindgren, Nataliya Kos'Myna.

SABRE is a 3-year project (2014-2017) funded by Labex CominLabs. It involves 1 Inria/IRISA team (Hybrid) and 2 groups from TELECOM BREST engineering school. The goal of SABRE is to improve computational functionnalities and power of current real-time EEG processing pipelines. The project will investigate innovative EEG solution methods empowered and speeded-up by ad-hoc, transistor-level, implementations of their key algorithmic operations. A completely new family of fully-hardware-integrated, new computational EEG imaging methods will be developed that are expected to speed up the imaging process of an EEG device of several orders of magnitude in real case scenarios.

9.1.5. IRT b<>com

Participants: Bruno Arnaldi [contact], Valérie Gouranton, Maud Marchal.

b<>com is a French Institute of Research and Technology (IRT). The main goal of this IRT is to fasten the development and marketing of tools, products and services in the field of digital technologies. Our team has collaborated with b<>com within two 3-year projects: ImData (on "Immersive Interaction") and GestChir (on "Augmented Healthcare") which both ended in 2016. A new 3-year project "NeedleWare" (on "Augmented Healthcare") has been started on October 2016.

9.1.6. CNPAO Project

Participants: Valérie Gouranton [contact], Jean-Baptiste Barreau, Ronan Gaugne.

CNPAO ("Conservatoire Numérique du Patrimoine Archéologique de l'Ouest") is an on-going research project partially funded by the Université Européenne de Bretagne (UEB) and Université de Rennes 1. It involves IRISA/Hybrid and CReAAH. The main objectives are: (i) a sustainable and centralized archiving of 2D/3D data produced by the archaeological community, (ii) a free access to metadata, (iii) a secure access to data for the different actors involved in scientific projects, and (iv) the support and advice for these actors in the 3D data production and exploration through the latest digital technologies, modeling tools and virtual reality systems.
This work was done in collaboration with Quentin Petit (SED Inria Rennes).

9.1.7. Imag'In CNRS IRMA

Participants: Bruno Arnaldi, Jean-Baptiste Barreau, Ronan Gaugne, Valérie Gouranton [contact].

The IRMA project is an Imag'In project funded by CNRS which aims at developping innovative methodologies for research in the field of cultural heritage based on the combination of medical imaging technologies and interactive 3D technologies (virtual reality, augmented reality, haptics, additive manufacturing). It relies on close collaborations with the National Institute of Preventive Archaeological Research (Inrap), the Research Center Archaeology, and History Archéosciences (CReAAH UMR 6566) and the company Image ET. The developed tools are intended for cultural heritage professionals such as museums, curators, restorers, and archaeologists. We focus on a large number of archeological artefacts of different nature, and various time periods (Paleolithic, Mesolithic, and Iron Age Medieval) from all over France. We can notably mention the oldest human bones found in Brittany (clavicle Beg Er Vil), a funeral urn from Trebeurden (22), or a Bronze Cauldron from a burial of the Merovingian necropolis "Crassés Saint-Dizier" (51). This project involves a strong collaboration with Théophane Nicolas (Inrap/UMR Trajectoires), Quentin Petit (SED Inria Rennes), and Grégor Marchand (CNRS/UMR CReAAH).

9.2. National Initiatives

9.2.1. ANR MANDARIN

Participants: Adrien Girard, Anatole Lécuyer, Maud Marchal [contact].

MANDARIN ("MANipulation Dextre hAptique pour opéRations INdustrielles en RV") was a 4-year ANR project (2012-2016). MANDARIN partners were CEA-List (coordinator), Inria/Hybrid, UTC, Haption and Renault. It aimed at designing new hardware and software solutions to achieve natural and intuitive mono and bi-manual dextrous interactions, suitable for virtual environments. The objective of Hybrid in MANDARIN was to design novel multimodal 3D interaction techniques and metaphors allowing to deal with haptic gloves limitations (portability, under-actuation) and to assist the user in virtual reality applications requiring dexterous manipulation. The results were evaluated with a representative industrial application: the bi-manual manipulation of complex rigid objects and cables bundles.

9.2.2. ANR HOMO-TEXTILUS

Participants: Anatole Lécuyer [contact], Maud Marchal.

HOMO-TEXTILUS was a 4-year ANR project (2012-2016). Partners of the project were : Inria/Hybrid, CHART, LIP6, TOMORROW LAND, RCP and potential end-user is Hussein Chalayan fashion designer. The objective of HOMO TEXTILUS was to study what could be the next generation of smart and augmented clothes, and their influence and potential impact on behavior and habits of their users. The project was strongly oriented towards human science, with both user studies and sociological studies. The involvement of Hybrid team in the project consisted in studying the design of next-gen prototypes of clothes embedding novel kinds of sensors and actuators. These prototypes were used and tested in various experiments.

9.2.3. FUI Previz

Participants: Bruno Arnaldi [contact], Valérie Gouranton [contact].

Previz was a 3-year project (2013-2016) funded by the competitive cluster "Images et Réseaux". Previz involved 4 Academic partners (Hybrid/INSA Rennes, ENS Louis-Lumière, LIRIS, Gipsa-Lab) and 9 Industrial partners (Technicolor, Ubisoft, SolidAnim, loumasystem, Polymorph). Previz aimed at proposing new previsualization tools for movie directors. The goal of Hybrid in Previz was to introduce new interactions between real and virtual actors so that the actor's actions, no matter his/her real or virtual nature, impact both the real and the virtual environment. The project ended up with a new production pipeline in order to automatically adapt and synchronize the visual effects (VFX), in space and time, to the real performance of an actor.

9.2.4. Ilab CertiViBE

Participants: Anatole Lécuyer [contact], Jussi Tapio Lindgren, Charles Garraud, Jérôme Chabrol.

CertiViBE is a 2-year "Inria Innovation Lab" (2015-2017) funded by Inria for supporting the development of OpenViBE software, and notably its evolution in order to enable and fasten the medical transfer and the medical certification of products based on OpenViBE. This joint lab involves two partners: Hybrid and Mensia Technologies startup company. The project aims at setting up a quality environment, and developping a novel version of the software wich should comply with medical certification rules.

9.2.5. IPL BCI-LIFT

Participants: Anatole Lécuyer [contact], Jussi Tapio Lindgren [contact], Andéol Evain, Lorraine Perronnet, Nataliya Kos'Myna.

BCI-LIFT is a 4-year "Inria Project Lab" initiative (2015-2019) funded by Inria for supporting a national research effort on Brain-Computer Interfaces. This joint lab involves sevearl Inria teams: Hybrid, Potioc, Athena, Neurosys, Mjolnir, Demar; as well as external partners: INSERM-Lyon, and INSA Rouen. This project aims at improving several aspects of Brain-Computer Interfaces : learning and adaptation of BCI systems, user interfaces and feedback, training protocols, etc.

9.3. European Initiatives

9.3.1. FP7 & H2020 Projects

9.3.1.1. HAPPINESS

Title: HAptic Printed Patterned INtErfaces for Sensitive Surface

Programm: H2020

Duration: January 2015 - December 2017

Coordinator: CEA

Partners:

- Arkema France (France)
- Robert Bosch (Germany)
- Commissariat A L'Energie Atomique et Aux Energies Alternatives (France)
- Fundacion Gaiker (Spain)
- Integrated Systems Development S.A. (Greece)
- University of Glasgow (United Kingdom)
- Walter Pak SL (Spain)

Inria contact: Nicolas Roussel and Anatole Lécuyer

The Automotive HMI (Human Machine Interface) will soon undergo dramatic changes, with large plastic dashboards moving from the 'push-buttons' era to the 'tactile' era. User demand for aesthetically pleasing and seamless interfaces is ever increasing, with touch sensitive interfaces now commonplace. However, these touch interfaces come at the cost of haptic feedback, which raises concerns regarding the safety of eyeless interaction during driving. The HAPPINESS project intends to address these concerns through technological solutions, introducing new capabilities for haptic feedback on these interfaces. The main goal of the HAPPINESS project is to develop a smart conformable surface able to offer different tactile sensations via the development of a Haptic Thin and Organic Large Area Electronic technology (TOLAE), integrating sensing and feedback capabilities, focusing on user requirements and ergonomic designs. To this aim, by gathering all the value chain actors (materials, technology manufacturing, OEM integrator) for application within the automotive market, the HAPPINESS project will offer a new haptic Human-Machine Interface technology, integrating touch sensing and disruptive feedback capabilities directly into an automotive dashboard. Based on the consortium skills, the HAPPINESS project will demonstrate the integration of Electro-Active Polymers (EAP) in a matrix of mechanical actuators on plastic foils. The objectives are to fabricate these actuators with large area and cost effective printing technologies and to integrate them through plastic molding injection into a small-scale dashboard prototype. We will design, implement and evaluate new approaches to Human-Computer Interaction on a fully functional prototype that combines in packaging both sensors and actuator foils, driven by custom electronics, and accessible to end-users via software libraries, allowing for the reproduction of common and accepted sensations such as Roughness, Vibration and Relief. In this project, the role of Hybrid team is to design user studies on tactile perception, and study innovative usages of the technologies developed in HAPPINESS.

9.4. International Research Visitors

9.4.1. Visits of International Scientists

Michael Pereira (EPFL, Switzerland) visited Hybrid for a collaboration on Brain-Computer Interfaces and sports in January 2016.

9.4.2. Visits to International Teams

Ferran Argelaguet visited the Virtual Reality Lab (Pr. Bernd Frohlich) at the Bauhaus University at Weimar (Germany) in October/November 2016.

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific events organisation

10.1.1.1. General chair, scientific chair

- Bruno Arnaldi was Scientific Chair of "Journées de l'AFRV" 2016, Brest, France.
- Anatole Lécuyer was Program Chair of IEEE Virtual Reality 2016.
- Maud Marchal was Program Chair of IEEE Symposium on 3D User Interfaces 2016.

10.1.2. Scientific events selection

10.1.2.1. Member of the conference program committees

- Anatole Lécuyer was Member of the conference program committee of Eurohaptics 2016.
- Ferran Argelaguet was Member of the conference program committee of IEEE Symposium on 3D User Interfaces 2016, ACM Virtual Reality Software and Technology 2016, and ACM Symposium on Spatial User Interfaces 2016.
- Maud Marchal was Member of the best paper committee of "Journées Françaises de l'Informatique Graphique" 2016.
- Valérie Gouranton was Member of the program committee of 3DCVE Workshop 2016.

10.1.2.2. Reviewer

- Ferran Argelaguet was Reviewer for ACM CHI 2016, Eurohaptics 2016.
- Maud Marchal was Reviewer for ACM Siggraph 2016, IEEE Virtual Reality 2016, IEEE Symposium on 3D User Interfaces 2016, Eurohaptics 2016.
- Valérie Gouranton was Reviewer for IEEE Virtual Reality 2016.

10.1.3. Journal

10.1.3.1. Member of the editorial boards

- Anatole Lécuyer is Associate Editor of Frontiers in Virtual Environments, and Presence journals.
- Ferran Argelaguet is Review Editor of Frontiers in Virtual Environments.
- Maud Marchal is Associate Editor of Computer Graphics Forum, Review Editor of Frontiers in Virtual Environments, and Member of the Editorial Board of Revue Francophone d'Informatique Graphique.

10.1.3.2. Reviewer - Reviewing activities

- Ferran Argelaguet was Reviewer for ACM Transactions on Graphics, IEEE Transactions on Visualization and Computer Graphics, and The Visual Computer.
- Maud Marchal was Reviewer for IEEE Transactions on Visualization and Computer Graphics, IEEE Transactions on Haptics, and The Visual Computer.
- Valérie Gouranton was Reviewer for Revue d'Intelligence Artificielle.

10.1.4. Invited talks

- Ferran Argelaguet was invited at the Bauhaus University Weimar (November 16).

10.1.5. Leadership within the scientific community

- Bruno Arnaldi is Vice-President of AFRV (French Association for Virtual Reality).
- Valérie Gouranton is Member of Executive Committee of AFRV (French Association for Virtual Reality).

10.1.6. Scientific expertise

- Bruno Arnaldi was Expert for French ANR (Agence Nationale de la Recherche).
- Maud Marchal was Expert for the Natural Sciences and Engineering Research Council of Canada (NSERC).

10.1.7. Research administration

- Bruno Arnaldi is Deputy Director of IRISA.
- Maud Marchal is Co-Head of the MRI Master Research in Computer Science.

10.2. Teaching - Supervision - Juries

Anatole Lécuyer:

Master MNRV: "Haptic Interaction", 9h, M2, ENSAM, Laval, FR

Ecole Centrale de Nantes : “Haptic Interaction and Brain-Computer Interfaces”, 4.5h, M1-M2, Ecole Centrale de Nantes, FR

Master SIBM: "Haptic and Brain-Computer Interfaces", 4.5h, M2, University of Rennes 1, FR

Bruno Arnaldi:

Master INSA Rennes: "VAR: Virtual and Augmented Reality", 12h, M2, INSA Rennes, FR

Master INSA Rennes: "Virtual Reality", courses 6h, projects 16h, M1 and M2, INSA Rennes, FR

Master INSA Rennes: Projects on "Virtual Reality", 20h, M1, INSA Rennes, FR

Ferran Argelaguet:

Master STS Informatique MITIC: "Techniques d'Interaction Avancées", 26h, M2, ISTIC, University of Rennes 1, FR

Master INSA Rennes: "Modeling and Engineering for Biology and Health Applications", 12h, M2, INSA Rennes, FR

Maud Marchal:

Master INSA Rennes: "Modeling and Engineering for Biology and Health Applications", 48h, M2 and responsible of this lecture, INSA Rennes, FR

Master SIBM: "Biomedical simulation", 3h, M2, University of Rennes 1, FR

Valérie Gouranton:

Licence: "Introduction to Virtual Reality", 22h, L2 and responsible of this lecture, INSA Rennes, FR

Licence: Project on "Virtual Reality", 16h, L3 and responsible of this lecture, INSA Rennes, FR

Master INSA Rennes: "Virtual Reality", 16h, M2, INSA Rennes, FR

Master INSA Rennes: Projects on "Virtual Reality", 20h, M1, INSA Rennes, FR

Florian Nouviale:

Ecole Centrale de Nantes: "Training on Unity3D", 6h, M1/M2, Ecole Centrale de Nantes, FR

Ronan Gaugne:

Insa Rennes: Projects on "Virtual Reality", 50h, L3/M1/M2, Insa Rennes, FR

10.2.1. Supervision

10.2.1.1. PhD (defended)

- Guillaume Claude, "The sequencing of actions in collaborative virtual environments", INSA Rennes, July 2016, Supervised by Bruno Arnaldi, Valérie Gouranton
- François Lehericey, "Détection de collision par lancer de rayon : la quête de la performance", INSA Rennes, September 2016, Supervised by Bruno Arnaldi, Valérie Gouranton
- Jérémy Lacoche, "Plasticity for User Interfaces in Mixed Reality", Université Rennes 1, July 2016, Supervised by Thierry Duval, Bruno Arnaldi, Jérôme Royan, Eric Maisel
- Morgan Le Chénéchal, "Awareness Model for Asymmetric Remote Collaboration in Mixed Reality", INSA Rennes, July 2016, Supervised by Bruno Arnaldi, Thierry Duval, Valérie Gouranton, Jérôme Royan
- Lucas Royer, "Visualization tools for needle insertion in interventional radiology", INSA Rennes, December 6th, 2016, Supervised by Alexandre Krupa and Maud aud Marchal
- Andéol Evain, "Optimizing the Use of SSVEP-based Brain-Computer Interfaces for Human-Computer Interaction", University of Rennes 1, December 6th, 2016, Supervised by Anatole Lécuyer, Nicolas Roussel, Géry Casiez and Ferran Argelaguet

10.2.1.2. PhD (in progress)

- Jean-Baptiste Barreau, "Virtual Reality and Archaelogy", Started in February 2014, Supervised by Valérie Gouranton and Bruno Arnaldi
- Benoit Le Gouis, "Multi-scale physical simulation", Started in October 2014, Supervised by Bruno Arnaldi, Maud Marchal and Anatole Lécuyer
- Lorraine Perronet, "Neurofeedback applications based on EEG, fMRI and VR", Started in January 2014, Supervised by Christian Barillot and Anatole Lécuyer
- Gwendal Le Moulec, "Automatic generation of VR applications", Started in October 2015, Supervised by Valérie Gouranton, Bruno Arnaldi and Arnaud Blouin
- Anne-Solène Dris-Kerdreux, "Training in virtual reality for construction applications", Started in November 2015, Supervised by Valérie Gouranton and Bruno Arnaldi
- Antoine Costes, "Haptic texturing", Started in November 2015, Supervised by Anatole Lécuyer and Ferran Argelaguet

- Guillaume Cortes, "Improving tracking in VR", Started in November 2015, Supervised by Anatole Lécuyer
- Hakim Si-Mohammed, "BCI and HCI", Started in October 2016, Supervised by Anatole Lécuyer and Ferran Argelaguet
- Gautier Picard, "Collaborative VR", Started in October 2016, Supervised by Valérie Gouranton, Bernard Gibaud and Bruno Arnaldi
- Hadrien Gurnel, "Prise en compte de la déformation d'organe pour l'assistance robotisée d'insertion d'aiguille", Started in October 2016, Supervised by Alexander Krupa and Maud Marchal

10.2.2. Juries

10.2.2.1. Selection committees

- Bruno Arnaldi was Member of Selection committee of Assistant Professor Position at Ecole Centrale de Nantes.

10.2.2.2. PhD and HDR juries

- Anatole Lécuyer was Member of PhD committees of Henrique Debarba (EPFL, Switzerland), Axelle Pillain (Telecom Bretagne), Maxence Rangé (Univ. Rennes 1), Jérémie Plouzeau (ENSAM Chalon), Andéol Evain (Univ. Rennes 1), and Member of HDR committee of Reinhold Scherer (TU Graz, Austria).
- Bruno Arnaldi was President of PhD committees of Kévin Jordao (INSA de Rennes), Hui-Yin Wu (Univ. Rennes 1), Charlotte Hoareau (Enib Brest), Sabhi Ahmed (Paris 8), Guillaume Claude (INSA Rennes), Morgan Le Chénéchal (INSA Rennes) and François Lehericey (INSA Rennes), Jérémy Lacoche (Univ. Rennes 1)
- Valérie Gouranton was Member of PhD committees of Guillaume Claude (INSA Rennes), Morgan Le Chénéchal (INSA Rennes) and François Lehericey (INSA Rennes)
- Maud Marchal was Reviewer of PhD theses of Johan Sarrazin (Univ. Grenoble Alpes), Pierre-Luc Manteaux (Univ. Grenoble Alpes) and Camille Schreck (Univ. Grenoble Alpes), and Member of PhD committee of Lucas Royer (INSA Rennes).
- Ferran Arguelaguet was Member of PhD committee of Andéol Evain (Univ. Rennes 1).

10.3. Popularization

The team has organized, together with MimeTIC team, a press conference and a press release on the "6-Finger Illusion" in May 2016. This event has been followed by lots of articles (internet, press) and radio coverages, including: France Inter, RFI, Europe 1, Le Monde, Libération, Les Echos, etc.

In addition, the results of the team have been disseminated in several other media coverages in 2016:

- "Journal télévisé 20h" (National prime time news), France2 channel (02/16) : presentation of the BCI activity.
- "Journal télévisé du soir", TVRennes channel (04/16) : presentation of the Forum Eurocities visit in Immersia.
- "La tête au carré", France Inter radio (10/16) : participation of Anatole Lécuyer.

The team has also participated to numerous dissemination events in 2016 (chronological order):

- "Open House ISTIC 2016" (Rennes, 01/16) : booth and demos of the team.
- "Forum EuroCities" (Rennes, 04/16) : presentation from Ronan Gaugne and demos in Immersia.
- "Les 10 ans du Quai Branly" (Paris, "Quai Branly" Museum, 06/16) : talk from Valérie Gouranton and Ronan Gaugne on virtual archaeology.
- "French-American Doctorat Exchange program 2016" (Rennes, 07/16): presentation from Ferran Argelaguet on Hybrid activities.
- "Journées du Patrimoine 2016" (Rennes, "Champs Libres" Museum, 09/16) : demos related to virtual archaelogy.
- "Journées Science et Musique 2016" (Rennes, 10/16) : co-organization of this event, and presentation of several demos.
- "Nuit Art et Science 2016" (Brest, 10/16) : demo on musical composition history in virtual reality and Immersia.
- "NEUROPLANETE 2016" (Nice, 10/16) : talk from Anatole Lécuyer and dissemination in "Le Point" journal.
- "Rencontres Inria Industrie: Interactions avec les objets et services numériques" (Lille, 11/16): demonstration from Ferran Argelaguet.

11. Bibliography

Publications of the year

Doctoral Dissertations and Habilitation Theses

[1] M. LE CHENECHAL.*Awareness Model for Asymmetric Remote Collaboration in Mixed Reality*, INSA de Rennes, July 2016, https://tel.archives-ouvertes.fr/tel-01392708.

Articles in International Peer-Reviewed Journal

[2] M. ACHIBET, A. GIRARD, M. MARCHAL, A. LÉCUYER.*Leveraging Passive Haptic Feedback in Virtual Environments with the Elastic-Arm Approach*, in "Presence: Teleoperators and Virtual Environments", 2016, vol. 25, n^{o} 1, p. 17 - 32 [*DOI :* 10.1162/PRES_A_00243], https://hal.inria.fr/hal-01414830.

[3] G. BRUDER, F. ARGELAGUET, A.-H. OLIVIER, A. LÉCUYER.*CAVE Size Matters: Effects of Screen Distance and Parallax on Distance Estimation in Large Immersive Display Setups*, in "Presence: Teleoperators and Virtual Environments", 2016, vol. 25, n^{o} 1, p. 1 - 16 [*DOI :* 10.1162/PRES_A_00241], https://hal.inria.fr/hal-01388499.

[4] G. CLAUDE, V. GOURANTON, B. CAILLAUD, B. GIBAUD, B. ARNALDI, P. JANNIN.*Synthesis and Simulation of Surgical Process Models*, in "Studies in Health Technology and Informatics", 2016, vol. 220, p. 63–70 [*DOI :* 10.3233/978-1-61499-625-5-63], https://hal.archives-ouvertes.fr/hal-01300990.

[5] A. GIRARD, M. MARCHAL, F. GOSSELIN, A. CHABRIER, F. LOUVEAU, A. LÉCUYER.*HapTip: Displaying Haptic Shear Forces at the Fingertips for Multi-Finger Interaction in Virtual Environments*, in "Frontiers in Robotics and AI", 2016 [*DOI :* 10.3389/FICT.2016.00006], https://hal.inria.fr/hal-01414862.

[6] L. HOYET, F. ARGELAGUET, C. NICOLE, A. LÉCUYER.*"Wow! I Have Six Fingers!": Would You Accept Structural Changes of Your Hand in VR?*, in "Frontiers in Robotics and AI", March 2016, vol. 3, n^{o} 27 [*DOI :* 10.3389/FROBT.2016.00027], https://hal.inria.fr/hal-01334359.

[7] V. KARAKKAT NARAYANAN, F. PASTEAU, M. MARCHAL, A. KRUPA, M. BABEL.*Vision-based adaptive assistance and haptic guidance for safe wheelchair corridor following*, in "Computer Vision and Image Understanding", August 2016, vol. 179, p. 171-185, https://hal.inria.fr/hal-01277585.

[8] L. ROYER, A. KRUPA, G. DARDENNE, A. LE BRAS, É. MARCHAND, M. MARCHAL.*Real-time Target Tracking of Soft Tissues in 3D Ultrasound Images Based on Robust Visual Information and Mechanical Simulation*, in "Medical Image Analysis", January 2017, vol. 35, p. 582 - 598 [*DOI :* 10.1016/J.MEDIA.2016.09.004], https://hal.inria.fr/hal-01374589.

[9] A. ÉVAIN, F. ARGELAGUET, G. CASIEZ, N. ROUSSEL, A. LÉCUYER.*Design and Evaluation of Fusion Approach for Combining Brain and Gaze Inputs for Target Selection*, in "Frontiers in Neuroscience", October 2016, vol. 10, n^{o} 454, 14 [*DOI :* 10.3389/FNINS.2016.00454], https://hal.inria.fr/hal-01388528.

[10] A. ÉVAIN, F. ARGELAGUET, G. CASIEZ, N. ROUSSEL, A. LÉCUYER.*Do the stimuli of an SSVEP-based BCI really have to be the same as the stimuli used for training it?*, in "Brain-Computer Interfaces", July 2016, vol. 3, n^{o} 2, p. 103 - 111 [*DOI :* 10.1080/2326263X.2016.1193458], https://hal.inria.fr/hal-01388534.

Articles in National Peer-Reviewed Journal

[11] T. NICOLAS, R. GAUGNE, C. TAVERNIER, V. GOURANTON, B. ARNALDI.*La tomographie, l'impression 3D et la réalité virtuelle au service de l'archéologie*, in "Les Nouvelles de l'archéologie", 2016, n^{o} 146, p. 16-22, https://hal.archives-ouvertes.fr/hal-01417753.

International Conferences with Proceedings

[12] M. ACHIBET, G. CASIEZ, M. MARCHAL.*DesktopGlove: a Multi-finger Force Feedback Interface Separating Degrees of Freedom Between Hands*, in "3DUI'16, the 11th Symposium on 3D User Interfaces", Greenville, United States, In proceedings of 3DUI'16, the 11th Symposium on 3D User Interfaces, IEEE Computer Society, March 2016, p. 3-12 [*DOI :* 10.1109/3DUI.2016.7460024], https://hal.inria.fr/hal-01267645.

[13] G. ANDRADE, F. NOUVIALE, J. ARDOUIN, É. MARCHAND, M. MARCHAL, A. LÉCUYER.*Enjoy 360° Vision with the FlyVIZ*, in "SIGGRAPH 2016 Emerging Technologies", Anaheim, United States, July 2016 [*DOI :* 10.1145/2929464.2929471], https://hal.inria.fr/hal-01387573.

[14] F. ARGELAGUET, L. HOYET, M. TRICO, A. LÉCUYER.*The role of interaction in virtual embodiment: Effects of the virtual hand representation*, in "IEEE Virtual Reality", Greenville, United States, March 2016, p. 3-10 [*DOI :* 10.1109/VR.2016.7504682], https://hal.inria.fr/hal-01346229.

[15] F. ARGELAGUET SANZ, M. MORGAN.*GiAnt: stereoscopic-compliant multi-scale navigation in VEs*, in "ACM Conference on Virtual Reality Software and Technology", Munich, Germany, November 2016, p. 269-277 [*DOI :* 10.1145/2993369.2993391], https://hal.inria.fr/hal-01393243.

[16] Y. BERNARD, J.-B. BARREAU, C. BIZIEN-JAGLIN, L. QUESNEL, L. LANGOUËT, M.-Y. DAIRE.*3d digitisation and reconstruction of a capital in northwestern gaul: interim results on the city of alet* , in "8th International Congress on Archaeology, Computer Graphics, Cultural Heritage and Innovation 'ARQUEOLÓGICA 2.0'", Valencia, Spain, Universitat Politècnica De Valencia, September 2016, p. 438-440, https://hal-insu.archives-ouvertes.fr/insu-01399040.

[17] R. BOUVILLE, V. GOURANTON, B. ARNALDI.*Virtual Reality Rehearsals for Acting with Visual Effects*, in "International Conference on Computer Graphics & Interactive Techniques", Victoria-BC, Canada, GI, 2016, p. 1-8, https://hal.inria.fr/hal-01314839.

[18] G. CLAUDE, V. GOURANTON, B. CAILLAUD, B. GIBAUD, P. JANNIN, B. ARNALDI.*From Observations to Collaborative Simulation: Application to Surgical Training*, in "ICAT-EGVE 2016 - International Conference on Artificial Reality and Telexistence, Eurographics Symposium on Virtual Environments", Little Rock, Arkansas, United States, December 2016, https://hal.archives-ouvertes.fr/hal-01391776.

[19] A.-S. DRIS, V. GOURANTON, B. ARNALDI.*Integration concept and model of Industry Foundation Classes (IFC) for interactive virtual environments*, in "33rd CIB W78 Conference", Brisbane, Australia, 2016, https://hal.archives-ouvertes.fr/hal-01391786.

[20] Y. GAFFARY, M. MARCHAL, A. GIRARD, M. PELLAN, A. ASSELIN, B. PEIGNÉ, M. EMILY, F. GOSSELIN, A. CHABRIER, A. LÉCUYER.*Studying One and Two-Finger Perception of Tactile Directional Cues*, in "10th International Conference on Haptics - Perception, Devices, Control, and Applications (EuroHaptics)", Londres, United Kingdom, F. BELLO, H. KAJIMOTO, Y. VISELL (editors), Haptics: Perception, Devices, Control, and Applications. 10th International Conference, EuroHaptics 2016, London, UK, July 4-7, 2016, Proceedings, Part II, Springer, July 2016, vol. 9775, p. 396-405, ISBN: 978-3-319-42324-1 ; 978-3-319-42323-4 [*DOI :* 10.1007/978-3-319-42324-1_39], https://hal.archives-ouvertes.fr/hal-01406434.

[21] J. LACOCHE, T. DUVAL, B. ARNALDI, É. MAISEL, J. ROYAN.*D3PART: A new Model for Redistribution and Plasticity of 3D User Interfaces*, in "3DUI 2016 : IEEE symposium on 3D User Interfaces Summit", Greenville, SC, United States, IEEE, March 2016, p. 23-36 [*DOI :* 10.1109/3DUI.2016.7460026], https://hal.archives-ouvertes.fr/hal-01293037.

[22] M. LE CHÉNÉCHAL, T. DUVAL, V. GOURANTON, J. ROYAN, B. ARNALDI.*Vishnu: Virtual Immersive Support for HelpiNg Users - An Interaction Paradigm for Collaborative Remote Guiding in Mixed Reality*, in "3DCVE 2016 : International Workshop on Collaborative Virtual Environments", Greenville, United States, IEEE, March 2016, p. 1 - 5 [*DOI :* 10.1109/3DCVE.2016.7563559], https://hal.archives-ouvertes.fr/hal-01293435.

[23] M. LE CHÉNÉCHAL, J. LACOCHE, J. ROYAN, T. DUVAL, V. GOURANTON, B. ARNALDI.*When the Giant meets the Ant An Asymmetric Approach for Collaborative and Concurrent Object Manipulation in a Multi-Scale Environment*, in "3DCVE 2016 : International Workshop on Collaborative Virtual Environments", Greenville, United States, IEEE, March 2016, p. 1 - 4 [*DOI :* 10.1109/3DCVE.2016.7563562], https://hal.archives-ouvertes.fr/hal-01293041.

[24] G. LE MOULEC, F. ARGELAGUET SANZ, A. LÉCUYER, V. GOURANTON.*Take-over control paradigms in collaborative virtual environments for training*, in "ACM Symposium on Virtual Reality Software and Technology (VRST)", Munich, Germany, 2016, https://hal.archives-ouvertes.fr/hal-01393003.

[25] T. NICOLAS, R. GAUGNE, C. TAVERNIER, V. GOURANTON, B. ARNALDI.*Internal 3D Printing of Intricate Structures*, in "6th International Conference on Culturage Heritage - EuroMed 2016", Nicosia, Cyprus, M. IOANNIDES, E. FINK, A. MOROPOULOU, M. HAGEDORN-SAUPE, A. FRESA, G. LIESTØL, V. RAJCIC, P. GRUSSENMEYER (editors), Lecture Notes in Computer Science, October 2016, vol. 10058, n° Part I, p. 432-441, https://hal.archives-ouvertes.fr/hal-01391762.

[26] E. SAFA, J.-B. BARREAU, R. GAUGNE, W. DUCHEMIN, J.-D. TALMA, B. ARNALDI, G. DUMONT, V. GOURANTON. *Digital and handcrafting processes applied to sound-studies of archaeological bone flutes*, in "International Conference on Culturage Heritage, EuroMed", Nicosia, Cyprus, 2016, vol. 1, n^{o} 10058, p. 184-195, https://hal.archives-ouvertes.fr/hal-01391755.

[27] A. ÉVAIN, F. ARGELAGUET SANZ, A. STROCK, N. ROUSSEL, G. CASIEZ, A. LÉCUYER. *Influence of Error Rate on Frustration of BCI Users*, in "AVI'16 - International Working Conference on Advanced Visual Interfaces", Bari, Italy, ACM, June 2016, p. 248–251 [*DOI :* 10.1145/2909132.2909278], https://hal.inria.fr/hal-01388552.

Scientific Books (or Scientific Book chapters)

[28] J. LINDGREN, A. LÉCUYER. *Introduction to BCI Software Platforms*, in "Brain-Computer Interfaces 2: Technology and Applications", M. CLERC, L. BOUGRAIN, F. LOTTE (editors), Wiley, August 2016, vol. 2, https://hal.inria.fr/hal-01420642.

[29] J. LINDGREN, A. LÉCUYER. *OpenViBE et les outils logiciels pour les BCI*, in "Les interfaces Cerveau-Ordinateur 2 : Technologies et Applications", M. CLERC, L. BOUGRAIN, F. LOTTE (editors), ISTE, July 2016, vol. 2, 336, https://hal.inria.fr/hal-01420637.

[30] A. LÉCUYER. *BCIs and Video Games: State of the Art with the OpenViBE2 Project*, in "Brain-Computer Interfaces 2: Technology and Applications", M. CLERC, L. BOUGRAIN, F. LOTTE (editors), Wiley, August 2016, vol. 1, https://hal.inria.fr/hal-01420623.

[31] A. LÉCUYER. *BCI et jeux vidéo : état de l'art à travers le projet OpenViBE2*, in "Les interfaces cerveau-ordinateur 2: Technologie et applications", M. CLERC, L. BOUGRAIN, F. LOTTE (editors), ISTE, July 2016, vol. 2, 336, https://hal.inria.fr/hal-01420629.

[32] L. PERRONNET, A. LÉCUYER, F. LOTTE, M. CLERC, C. BARILLOT. *Brain training with neurofeedback*, in "Brain-Computer Interfaces 1", Wiley-ISTE, July 2016, https://hal.inria.fr/hal-01413424.

[33] L. PERRONNET, A. LÉCUYER, F. LOTTE, M. CLERC, C. BARILLOT. *Entraîner son cerveau avec le neurofeedback*, in "Les interfaces cerveau-ordinateur 1", M. CLERC, L. BOUGRAIN, F. LOTTE (editors), July 2016, https://hal.inria.fr/hal-01413408.

[34] A. ÉVAIN, N. ROUSSEL, G. CASIEZ, F. ARGELAGUET SANZ, A. LÉCUYER. *Brain-Computer Interfaces for Human–Computer Interaction*, in "Brain-Computer Interfaces 1: Foundations and Methods", M. CLERC, L. BOUGRAIN, F. LOTTE (editors), Wiley-ISTE, August 2016, https://hal.inria.fr/hal-01416382.

Other Publications

[35] M. LE CHÉNÉCHAL, J. LACOCHE, J. ROYAN, T. DUVAL, V. GOURANTON, B. ARNALDI. *When the Giant meets the Ant An Asymmetric Approach for Collaborative Object Manipulation*, IEEE, March 2016, p. 273 - 274, 3DUI 2016 : 11th IEEE Symposium on 3D User Interfaces, Poster [*DOI :* 10.1109/3DUI.2016.7460078], https://hal.archives-ouvertes.fr/hal-01301770.

References in notes

[36] M. CLERC, L. BOUGRAIN, F. LOTTE (editors). *Les interfaces Cerveau-Ordinateur 1*, ISTE, July 2016, https://hal.inria.fr/hal-01402539.

[37] M. CLERC, L. BOUGRAIN, F. LOTTE (editors). *Les interfaces cerveau-ordinateur 2*, ISTE, July 2016, https://hal.inria.fr/hal-01402544.

[38] D. A. BOWMAN, E. KRUIJFF, J. J. LAVIOLA, I. POUPYREV.*3D User Interfaces: Theory and Practice*, Addison Wesley, 2004.

[39] M. CLERC, L. BOUGRAIN, F. LOTTE. , F. L. MAUREEN CLERC (editor)*Brain-Computer Interfaces 1*, Wiley-ISTE, July 2016, https://hal.inria.fr/hal-01408991.

[40] M. CLERC, L. BOUGRAIN, F. LOTTE. , F. L. MAUREEN CLERC (editor)*Brain-Computer Interfaces 2*, Wiley-ISTE, July 2016, https://hal.inria.fr/hal-01408998.

[41] A. LÉCUYER.*Simulating Haptic Feedback Using Vision: A Survey of Research and Applications of Pseudo-Haptic Feedback*, in "Presence: Teleoperators and Virtual Environments", January 2009, vol. 18, n^{o} 1, p. 39–53, http://www.mitpressjournals.org/doi/abs/10.1162/pres.18.1.39.

Project-Team HYCOMES

Modélisation hybride & conception par contrats pour les systèmes embarqués multi-physiques

RESEARCH CENTER
Rennes - Bretagne-Atlantique

THEME
Embedded and Real-time Systems

Table of contents

Project-Team HYCOMES

Creation of the Team: 2013 July 01, updated into Project-Team: 2016 September 01

Keywords:

Computer Science and Digital Science:

2. - Software
2.1. - Programming Languages
2.1.1. - Semantics of programming languages
2.1.5. - Constraint programming
2.1.8. - Synchronous languages
2.1.10. - Domain-specific languages
2.2. - Compilation
2.3. - Embedded and cyber-physical systems
2.3.1. - Embedded systems
2.3.2. - Cyber-physical systems
2.3.3. - Real-time systems
2.4. - Verification, reliability, certification
2.4.1. - Analysis
2.4.2. - Model-checking
2.4.3. - Proofs
2.5. - Software engineering
2.5.1. - Software Architecture & Design
2.5.2. - Component-based Design
3. - Data and knowledge
3.1. - Data
3.1.1. - Modeling, representation
6. - Modeling, simulation and control
6.1. - Mathematical Modeling
6.1.1. - Continuous Modeling (PDE, ODE)
6.1.3. - Discrete Modeling (multi-agent, people centered)
6.1.5. - Multiphysics modeling

Other Research Topics and Application Domains:

2. - Health
2.4. - Therapies
2.4.3. - Surgery
4. - Energy
5. - Industry of the future
5.2. - Design and manufacturing
5.2.1. - Road vehicles
5.2.2. - Railway
5.2.3. - Aviation
5.2.4. - Aerospace

5.8. - Learning and training
5.9. - Industrial maintenance
7. - Transport and logistics
7.1. - Traffic management
7.1.3. - Air traffic
8. - Smart Cities and Territories
8.1. - Smart building/home
8.1.1. - Energy for smart buildings

1. Members

Research Scientists

Benoît Caillaud [Team leader, Inria, Senior Researcher, HDR]
Albert Benveniste [Inria, Senior Researcher, Emeritus, HDR]
Khalil Ghorbal [Inria, Researcher]

Technical Staff

Aurélien Lamercerie [Inria, from Nov 2016, funded by the Sunset collaborative project, Labex CominLabs]

PhD Student

Ayman Aljarbouh [Inria, partially granted by Conseil Régional de Bretagne and by the Modrio ITEA2 project]

Administrative Assistant

Angélique Jarnoux [Inria]

Other

Siham Rim Boudaoud [Univ. Rennes I, Master Intern, from Feb 2016 until Jun 2016]

2. Overall Objectives

2.1. Overall Objectives

Hycomes has been created as a new team of the Rennes — Bretagne Atlantique Inria research center in July 2013. The team builds upon the most promising results of the former S4 team-project and of the Synchronics large scale initiative. Two topics in embedded system design are covered:

- Hybrid systems modelling, with applications to the design of multi-physics embedded systems, often referenced as cyber-physical systems;
- Contract-based design and interface theories, with applications to requirements engineering in the context of safety-critical systems design.

3. Research Program

3.1. Hybrid Systems Modeling

Systems industries today make extensive use of mathematical modeling tools to design computer controlled physical systems. This class of tools addresses the modeling of physical systems with models that are simpler than usual scientific computing problems by using only Ordinary Differential Equations (ODE) and Difference Equations but not Partial Differential Equations (PDE). This family of tools first emerged in the 1980's with SystemBuild by MatrixX (now distributed by National Instruments) followed soon by Simulink by Mathworks, with an impressive subsequent development.

In the early 90's control scientists from the University of Lund (Sweden) realized that the above approach did not support component based modeling of physical systems with reuse [0]. For instance, it was not easy to draw an electrical or hydraulic circuit by assembling component models of the various devices. The development of the Omola language by Hilding Elmqvist was a first attempt to bridge this gap by supporting some form of Differential Algebraic Equations (DAE) in the models. Modelica quickly emerged from this first attempt and became in the 2000's a major international concerted effort with the Modelica Consortium [0]. A wider set of tools, both industrial and academic, now exists in this segment [0]. In the EDA sector, VHDL-AMS was developed as a standard [13].

Despite these tools are now widely used by a number of engineers, they raise a number of technical difficulties. The meaning of some programs, their mathematical semantics, can be tainted with uncertainty. A main source of difficulty lies in the failure to properly handle the discrete and the continuous parts of systems, and their interaction. How the propagation of mode changes and resets should be handled? How to avoid artifacts due to the use of a global ODE solver causing unwanted coupling between seemingly non interacting subsystems? Also, the mixed use of an equational style for the continuous dynamics with an imperative style for the mode changes and resets is a source of difficulty when handling parallel composition. It is therefore not uncommon that tools return complex warnings for programs with many different suggested hints for fixing them. Yet, these "pathological" programs can still be executed, if wanted so, giving surprising results — See for instance the Simulink examples in [21], [1] and [17].

Indeed this area suffers from the same difficulties that led to the development of the theory of synchronous languages as an effort to fix obscure compilation schemes for discrete time equation based languages in the 1980's. Our vision is that hybrid systems modeling tools deserve similar efforts in theory as synchronous languages did for the programming of embedded systems.

3.2. Background on non-standard analysis

Non-Standard analysis plays a central role in our research on hybrid systems modeling [1], [21], [18], [17]. The following text provides a brief summary of this theory and gives some hints on its usefulness in the context of hybrid systems modeling. This presentation is based on our paper [1], a chapter of Simon Bliudze's PhD thesis [27], and a recent presentation of non-standard analysis, not axiomatic in style, due to the mathematician Lindström [50].

Non-standard numbers allowed us to reconsider the semantics of hybrid systems and propose a radical alternative to the *super-dense time semantics* developed by Edward Lee and his team as part of the Ptolemy II project, where cascades of successive instants can occur in zero time by using $\mathbb{R}_+ \times \mathbb{N}$ as a time index. In the non-standard semantics, the time index is defined as a set $\mathbb{T} = \{n\partial \mid n \in {}^*\mathbb{N}\}$, where ∂ is an *infinitesimal* and ${}^*\mathbb{N}$ is the set of *non-standard integers*. Remark that 1/ $\mathbb{T}$ is dense in $\mathbb{R}_+$, making it "continuous", and 2/ every $t \in \mathbb{T}$ has a predecessor in $\mathbb{T}$ and a successor in $\mathbb{T}$, making it "discrete". Although it is not effective from a computability point of view, the *non-standard semantics* provides a framework that is familiar to the computer scientist and at the same time efficient as a symbolic abstraction. This makes it an excellent candidate for the development of provably correct compilation schemes and type systems for hybrid systems modeling languages.

Non-standard analysis was proposed by Abraham Robinson in the 1960s to allow the explicit manipulation of "infinitesimals" in analysis [56], [42], [12]. Robinson's approach is axiomatic; he proposes adding three new axioms to the basic Zermelo-Fraenkel (ZFC) framework. There has been much debate in the mathematical community as to whether it is worth considering non-standard analysis instead of staying with the traditional one. We do not enter this debate. The important thing for us is that non-standard analysis allows the use of the non-standard discretization of continuous dynamics "as if" it was operational.

[0] http://www.lccc.lth.se/media/LCCC2012/WorkshopSeptember/slides/Astrom.pdf
[0] https://www.modelica.org/
[0] SimScape by Mathworks, Amesim by LMS International, now Siemens PLM, and more.

Not surprisingly, such an idea is quite ancient. Iwasaki et al. [46] first proposed using non-standard analysis to discuss the nature of time in hybrid systems. Bliudze and Krob [28], [27] have also used non-standard analysis as a mathematical support for defining a system theory for hybrid systems. They discuss in detail the notion of "system" and investigate computability issues. The formalization they propose closely follows that of Turing machines, with a memory tape and a control mechanism.

The introduction to non-standard analysis in [27] is very pleasant and we take the liberty to borrow it. This presentation was originally due to Lindstrøm, see [50]. Its interest is that it does not require any fancy axiomatic material but only makes use of the axiom of choice — actually a weaker form of it. The proposed construction bears some resemblance to the construction of $\mathbb{R}$ as the set of equivalence classes of Cauchy sequences in $\mathbb{Q}$ modulo the equivalence relation $(u_n) \approx (v_n)$ iff $\lim_{n\to\infty} (u_n - v_n) = 0$.

3.3. Contract-Based Design, Interfaces Theories, and Requirements Engineering

System companies such as automotive and aeronautic companies are facing significant difficulties due to the exponentially raising complexity of their products coupled with increasingly tight demands on functionality, correctness, and time-to-market. The cost of being late to market or of imperfections in the products is staggering as witnessed by the recent recalls and delivery delays that many major car and airplane manufacturers had to bear in the recent years. The specific root causes of these design problems are complex and relate to a number of issues ranging from design processes and relationships with different departments of the same company and with suppliers, to incomplete requirement specification and testing.

We believe the most promising means to address the challenges in systems engineering is to employ structured and formal design methodologies that seamlessly and coherently combine the various viewpoints of the design space (behavior, space, time, energy, reliability, ...), that provide the appropriate abstractions to manage the inherent complexity, and that can provide correct-by-construction implementations. The following technology issues must be addressed when developing new approaches to the design of complex systems:

- The overall design flows for heterogeneous systems and the associated use of models across traditional boundaries are not well developed and understood. Relationships between different teams inside a same company, or between different stake-holders in the supplier chain, are not well supported by solid technical descriptions for the mutual obligations.
- System requirements capture and analysis is in large part a heuristic process, where the informal text and natural language-based techniques in use today are facing significant challenges. Formal requirements engineering is in its infancy: mathematical models, formal analysis techniques and links to system implementation must be developed.
- Dealing with variability, uncertainty, and life-cycle issues, such as extensibility of a product family, are not well-addressed using available systems engineering methodologies and tools.

The challenge is to address the entire process and not to consider only local solutions of methodology, tools, and models that ease part of the design.

Contract-based design has been proposed as a new approach to the system design problem that is rigorous and effective in dealing with the problems and challenges described before, and that, at the same time, does not require a radical change in the way industrial designers carry out their task as it cuts across design flows of different type. Indeed, contracts can be used almost everywhere and at nearly all stages of system design, from early requirements capture, to embedded computing infrastructure and detailed design involving circuits and other hardware. Contracts explicitly handle pairs of properties, respectively representing the assumptions on the environment and the guarantees of the system under these assumptions. Intuitively, a contract is a pair $C = (A, G)$ of assumptions and guarantees characterizing in a formal way 1) under which context the design is assumed to operate, and 2) what its obligations are. Assume/Guarantee reasoning has been known for a long time, and has been used mostly as verification mean for the design of software [54]. However, contract based design with explicit assumptions is a philosophy that should be followed all along the design, with all kinds of models, whenever necessary. Here, specifications are not limited to profiles, types, or taxonomy of data, but also describe the functions, performances of various kinds (time and energy), and reliability. This amounts to enrich a component's interface with, on one hand, formal specifications of the behavior of the environment in which the component may be instantiated and, on the other hand, of the expected behavior of the component itself. The consideration of rich interfaces is still in its infancy. So far, academic researchers have addressed the mathematics and algorithmics of interfaces theories and contract-based reasoning. To make them a technique of choice for system engineers, we must develop:

- Mathematical foundations for interfaces and requirements engineering that enable the design of frameworks and tools;
- A system engineering framework and associated methodologies and tool sets that focus on system requirements modeling, contract specification, and verification at multiple abstraction layers.

A detailed bibliography on contract and interface theories for embedded system design can be found in [2]. In a nutshell, contract and interface theories fall into two main categories:

Assume/guarantee contracts. By explicitly relying on the notions of assumptions and guarantees, A/G-contracts are intuitive, which makes them appealing for the engineer. In A/G-contracts, assumptions and guarantees are just properties regarding the behavior of a component and of its environment. The typical case is when these properties are formal languages or sets of traces, which includes the class of safety properties [47], [36], [53], [15], [37]. Contract theories were initially developed as specification formalisms able to refuse some inputs from the environment [43]. A/G-contracts were advocated by the SPEEDS project [20]. They were further experimented in the framework of the CESAR project [38], with the additional consideration of *weak* and *strong* assumptions. This is still a very active research topic, with several recent contributions dealing with the timed [25] and probabilistic [32], [33] viewpoints in system design, and even mixed-analog circuit design [55].

Automata theoretic interfaces. Interfaces combine assumptions and guarantees in a single, automata theoretic specification. Most interface theories are based on Lynch Input/Output Automata [52], [51]. Interface Automata [59], [58], [60], [34] focus primarily on parallel composition and compatibility: Two interfaces can be composed and are compatible if there is at least one environment where they can work together. The idea is that the resulting composition exposes as an interface the needed information to ensure that incompatible pairs of states cannot be reached. This can be achieved by using the possibility, for an Interface Automaton, to refuse selected inputs from the environment in a given state, which amounts to the implicit assumption that the environment will never produce any of the refused inputs, when the interface is in this state. Modal Interfaces [3] inherit from both Interface Automata and the originally unrelated notion of Modal Transition System [49], [14], [29], [48]. Modal Interfaces are strictly more expressive than Interface Automata by decoupling the I/O orientation of an event and its deontic modalities (mandatory, allowed or forbidden). Informally, a *must* transition is available in every component that realizes the modal interface, while a *may* transition needs not be. Research on interface theories is still very active. For instance, timed [61], [22], [24], [40], [39], [23], probabilistic [32], [41] and energy-aware [35] interface theories have been proposed recently.

Requirements Engineering is one of the major concerns in large systems industries today, particularly so in sectors where certification prevails [57]. DOORS projects collecting requirements are poorly structured and cannot be considered a formal modeling framework today. They are nothing more than an informal documentation enriched with hyperlinks. As examples, medium size sub-systems may have a few thousands requirements and the Rafale fighter aircraft has above 250,000 of them. For the Boeing 787, requirements were not stable while subcontractors performed the development of the fly-by-wire and of the landing gear subsystems.

We see Contract-Based Design and Interfaces Theories as innovative tools in support of Requirements Engineering. The Software Engineering community has extensively covered several aspects of Requirements Engineering, in particular:

- the development and use of large and rich *ontologies*; and

- the use of Model Driven Engineering technology for the structural aspects of requirements and resulting hyperlinks (to tests, documentation, PLM, architecture, and so on).

Behavioral models and properties, however, are not properly encompassed by the above approaches. This is the cause of a remaining gap between this phase of systems design and later phases where formal model based methods involving behavior have become prevalent—see the success of Matlab/Simulink/Scade technologies. We believe that our work on contract based design and interface theories is best suited to bridge this gap.

4. Application Domains

4.1. Cyber-Physical Systems Design

Academic research and industry are currently witnessing several major revolutions: *Cyber-Physical Systems* (CPS), *Big-Data* and *Cloud Computing*, just to name a few. The Hycomes team is focused on CPS, and more precisely on CPS modeling with two targeted applications: The rigorous design of CPS and the optimal exploitation of CPS. Despite many engineers believe that *systems become too complex to be modeled in a faithfully*, the Hycomes team defends the opposite idea. We believe in the benefits of modeling, but acknowledge that the communities of researchers and tool developers are in part responsible for this defiance. The steep increase in the complexity of systems (e.g., public transportation systems, electric power grids) and of their models comes from composing smaller subsystems into complex architectures. As a matter of fact, these architectures are sparse, and subsystems interactions are confined to immediate surrounding neighborhoods. Thus, the dimension (number of state variables) of a system is not the most appropriate characterization of its complexity. It is rather the structure of a system and its combinatorics of modes of operation that encapsulate its complexity.

The main objective of the Hycomes team is to advance modeling technologies (languages, compile-time analyses, simulation techniques) for CPS combining physical interactions, communication layers and software components. We believe that mastering CPS comprising thousands to millions of components requires radical changes of paradigms. For instance, modeling techniques must be revised, especially when physics is involved. Modeling languages must be enhanced to cope with larger models. This can only be done by combining new **compilation** techniques (to master the structural complexity of models) with new **mathematical** tools (new numerical methods, in particular). We identify below the different axis we want to tackle.

4.1.1. Modelica

Modelica is a component-based modeling language initially designed for the modeling of multi-physics systems. The mathematical paradigm underlying Modelica, known as *Differential Algebraic Equations* (DAE). The key challenge is to be able to combine algebraic constraints, resulting from the laws of physics, in interaction with the nonsmooth behavior of some physical phenomena (e.g., impact laws), the multiple modes of operation of the system, and the intrinsically discrete behavior of software components. In essence, Modelica is based on the concept of multi-mode DAE, so that models can switch from one behavior to another when an event occurs, typically the crossing of a threshold. This approach is paramount to the modeling of large CPS. For instance, EDF has done a thorough modeling of the electric power grid of the Reunion island [0]. This was undertaken to gain a better understanding of this complex and notably unstable assembly of highly decentralized electric power plants: dams, small thermal power plants, wind and solar farms, and residential solar panels, just to name a few. This large model turned out to be intractable with state-of-the-art Modelica tools: because Modelica compilation techniques are not modular, the whole model has to be compiled as one unit, resulting in a very large simulation code. Parallel simulation of Modelica models is still in its infancy and gives poor results on very large models [44]; parallel/distributed techniques for networks of FMU components are not applicable to a monolithic model [45], [16]. Moreover, when simulating, for instance, thermal models of a building, the opening of a window or of a door impacts the whole simulation, despite it only has a local impact on the heat exchanges and temperatures. This is caused by the sudden change of stiffness in some part

[0] http://www.ceser-reunion.fr/fileadmin/user_upload/tx_pubdb/archives/10.10.18_Rapport_electricite.pdf

of the model, that forces a change in discretization step size (assuming that a variable step solver is used for simulation), with the adverse effect that the simulation of the whole system is slowed down. The root cause of this phenomenon boils down to the fact that system models and numerical methods used to simulate them are not space adaptive — recall that such models are 0-D models, with ODEs/DAEs, with no Partial Differential Equations (PDEs).

4.1.2. Co-modeling and co-simulation

The emergence of the FMI standard [0] supporting co-modeling and co-simulation has contributed to the widespread belief that the co-simulation of a large number of models is achievable using FMI-based tools. This is unfortunately an illusion, as FMI does not guarantee the reproducibility and determinacy of simulations. There are several reasons for that. First, FMI offers no rollback mechanism [30], which makes the co-simulation to depend on the discretization policy. Second, as the standard is not formally specified, its various implementations by tool developers differ.

4.1.3. Beyond simulation

Many physical science engineers (mechanical, electrical, aeronautic, ...) develop models with the sole objective to simulate them, while it is known that models can be used for a variety of tasks, all contributing towards the safe design and operation of a CPS: validating a design model against a set of requirements, assess the robustness of a model, test implementations against a design model, perform state estimation during system operation, just to name a few.

Early stages of CPS design usually consist in the elicitation of system-level requirements that will be used later on to design detailed models that can be simulated. Most often, the design tasks are split among several suppliers. This calls for precise requirements to be passed to them, so that, as far as feasible, suppliers can work independently. Some of the requirements specify the allowed behavior of the sub-system to be design, while others specify the assumed behavior of the sub-system's environment.

During operation of a CPS, maintenance tasks play an ever-increasing role, to minimize the downtime of the system and, to maintain an extremely low probability of occurrence of catastrophic failures. *Diagnosis* enables to replace some routine inspections or precautionary replacements of critical parts (that are usually triggered by the number of hours of operation, or by calendar) by fewer maintenance operations, triggered by the estimated wear or aging of those parts. This helps to reduce immobilization times and maintenance costs. Design models could be reused to help the development of diagnosis software that will trigger maintenance operations, based on the output of *parity check* algorithms [26], capable of detecting slow or sudden changes of some parameters. Reusing design models in this context would be a genuine innovation, in comparison to the established practice, where diagnosis is designed by hand, from scratch.

4.1.4. Verification

Because of severe complexity or undecidability problems, CPS formal verification can be done only on partial and simplified models. When applicable, these techniques complement usefully simulations. Despite of the high level of expertise it requires, formal verification brings a level of confidence in the analyses that can not be compared with what can be obtained by simulation. Using formal verification makes sense only for the most critical parts of a CPS. A fine example is the formal correctness proof of a new generation of aircraft collision prevention system, the ACAS-X [6]. This proof has facilitated the certification of this system, according to the established aeronautic standards (DO-178C [0]).

5. Highlights of the Year

5.1. Highlights of the Year

[0] https://www.fmi-standard.org/
[0] http://www.adacore.com/gnatpro-safety-critical/avionics/do178c/

Team members have made a significant step towards the definition of a formal semantics of multimode DAE systems, their strucutral analysis and the generation of simulation code. In particular, impulsive behavior at mode changes are handled correctly [19] (see Section 7.1 for full details). This semantics has been implemented, in part, in the SunDAE prototype software (Section 6.1).

6. New Software and Platforms

6.1. SunDAE

Structural analysis tool for multimode DAE systems
FUNCTIONAL DESCRIPTION

SunDAE is a multimode DAE (mDAE) structural analysis tool. Structural differentiation index is determined, impulsion analysis is performed and a BTF scheduling of the equations is performed, for each mode of a mDAE system. The input language consists in guarded equations. The output is a state-machine where states define continuous-time dynamics and transitions define resets. Both are defined by scheduled blocks of equations. SunDAE has been developed since 2016 by the Hycomes team and is distributed as an open-source software, under the CeCCIL Free Software Licensing Agreement.

- Contact: Benoit Caillaud

6.2. Flipflop

Test & Flip Net Synthesis Tool for the Inference of Technical Procedure Models
FUNCTIONAL DESCRIPTION

Flipflop is a Test and Flip net synthesis tool implementing a linear algebraic polynomial time algorithm. Computations are done in the Z/2Z ring. Test and Flip nets extend Elementary Net Systems by allowing test to zero, test to one and flip arcs. The effect of flip arcs is to complement the marking of the place. While the net synthesis problem has been proved to be NP hard for Elementary Net Systems, thanks to flip arcs, the synthesis of Test and Flip nets can be done in polynomial time. Test and flip nets have the required expressivity to give concise and accurate representations of surgical processes (models of types of surgical operations). Test and Flip nets can express causality and conflict relations. The tool takes as input either standard XES log files (a standard XML file format for process mining tools) or a specific XML file format for surgical applications. The output is a Test and Flip net, solution of the following synthesis problem: Given a finite input language (log file), compute a net, which language is the least language in the class of Test and Flip net languages, containing the input language.

- Contact: Benoit Caillaud
- URL: http://tinyurl.com/oql6f3y

6.3. MICA

Model Interface Compositional Analysis Library
KEYWORDS: Modal interfaces - Contract-based desing
SCIENTIFIC DESCRIPTION

In Mica, systems and interfaces are represented by extension. However, a careful design of the state and event heap enables the definition, composition and analysis of reasonably large systems and interfaces. The heap stores states and events in a hash table and ensures structural equality (there is no duplication). Therefore complex data-structures for states and events induce a very low overhead, as checking equality is done in constant time.

Thanks to the Inter module and the mica interactive environment, users can define complex systems and interfaces using Ocaml syntax. It is even possible to define parameterized components as Ocaml functions.
FUNCTIONAL DESCRIPTION

Mica is an Ocaml library implementing the Modal Interface algebra. The purpose of Modal Interfaces is to provide a formal support to contract based design methods in the field of system engineering. Modal Interfaces enable compositional reasoning methods on I/O reactive systems.

- Participant: Benoit Caillaud
- Contact: Benoit Caillaud
- URL: http://www.irisa.fr/s4/tools/mica/

7. New Results

7.1. Structural Analysis of Multi-Mode DAEs

Differential Algebraic Equation (DAE) systems constitute the mathematical model supporting physical modeling languages such as Modelica or Simscape. Unlike Ordinary Differential Equations, or ODEs, they exhibit subtle issues because of their implicit *latent equations* and related *differentiation index*. Multi-mode DAE (mDAE) systems are much harder to deal with, not only because of their mode-dependent dynamics, but essentially because of the events and resets occurring at mode transitions. Unfortunately, the large literature devoted to the numerical analysis of DAEs do not cover the multi-mode case. It typically says nothing about mode changes. This lack of foundations cause numerous difficulties to the existing modeling tools. Some models are well handled, others are not, with no clear boundary between the two classes. In [11], we develop a comprehensive mathematical approach to the *structural analysis* of mDAE systems which properly extends the usual analysis of DAE systems. We define a constructive semantics based on nonstandard analysis and show how to produce execution schemes in a systematic way. This work has been accepted for presentation at the HSCC 2017 conference [19] in April 2017.

7.2. Decoupling Abstractions

In [10], we investigated decoupling abstractions, by which we seek to simulate (i.e. abstract) a given system of ordinary differential equations (ODEs) by another system that features completely independent (i.e. uncoupled) sub-systems, which can be considered as separate systems in their own right. Beyond a purely mathematical interest as a tool for the qualitative analysis of ODEs, decoupling can be applied to verification problems arising in the fields of control and hybrid systems. Existing verification technology often scales poorly with dimension. Thus, reducing a verification problem to a number of independent verification problems for systems of smaller dimension may enable one to prove properties that are otherwise seen as too difficult. We show an interesting correspondence between Darboux polynomials and decoupling simulating abstractions of systems of polynomial ODEs and give a constructive procedure for automatically computing the latter.

7.3. Formal Verification of the ACAS X System

The *Next-Generation Airborne Collision Avoidance System* (ACAS X) is intended to be installed on all large aircraft to give advice to pilots and prevent mid-air collisions with other aircraft. It is currently being developed by the Federal Aviation Administration (FAA). In [6], we determine the geometric configurations under which the advice given by ACAS X is safe under a precise set of assumptions and formally verify these configurations using hybrid systems theorem proving techniques. We consider subsequent advisories and show how to adapt our formal verification to take them into account. We examine the current version of the real ACAS X system and discuss some cases where our safety theorem conflicts with the actual advisory given by that version, demonstrating how formal hybrid systems proving approaches are helping to ensure the safety of ACAS X. Our approach is general and could also be used to identify unsafe advice issued by other collision avoidance systems or confirm their safety.

7.4. Chattering-Free Simulation

Chattering is a fundamental phenomenon that is unique to hybrid systems, due to the complex interaction between discrete dynamics (in the form of discrete transitions) and continuous dynamics (in the form of time). In practice, simulating chattering hybrid systems is challenging in that simulation effectively halts near the chattering time point, as an infinite number of discrete transitions would need to be simulated. In [7], formal conditions are provided for when the simulated models of hybrid systems display chattering behavior, and methods are proposed for avoiding chattering "on the fly" in runtime. We utilize dynamical behavior analysis to derive conditions for detecting chattering without enumeration of modes. We also present a new iterative algorithm to allow for solutions to be carried past the chattering point, and we show by a prototypical implementation how to generate the equivalent chattering-free dynamics internally by the simulator in the main simulation loop.

8. Partnerships and Cooperations

8.1. Regional Initiatives

- Ayman Aljarbouh's PhD is partially funded by an ARED grant of the Brittany Regional Council. His doctoral work took place in the context of the Modrio (completed in 2016) and Sys2Soft (completed in 2015) projects on hybrid systems modeling. Ayman Aljarbouh is working on accelerated simulation techniques for hybrid systems. In particular, he is focusing on the regularisation, at runtime, of chattering behaviour and the approximation of Zeno behaviour.
- Benoît Caillaud and Aurélien Lamercerie are participating to the S3PM and SUNSET projects of the CominLabs excellence laboratory [0]. This project focuses on the computation of surgical procedural knowledge models from recordings of individual procedures, and their execution [31]. The objective is to develop an enabling technology for procedural knowledge based computer assistance of surgery. In this project, we demonstrate its potential added value in nurse and surgeon training [9], [5].

8.2. European Initiatives

8.2.1. Collaborations in European Programs, Except FP7 & H2020

Program: ITEA2

Project acronym: Modrio

Project title: Model Driven Physical Systems Operation

Duration: September 2012 – May 2016

Coordinator: EDF (France)

Other partners: ABB (Sweden), Ampère Laboratory / CNRS (France), Bielefeld University (Germany), Dassault Systèmes (Sweden), Dassault Aviation (France), DLR (Germany), DPS (France), EADS (France), Equa Simulation (Sweden), IFP (France), ITI (Germany), Ilmenau University (Germany), Katholic University of Leuven (Belgium), Knorr-Bremse (Germany), LMS (France and Belgium), Linköping University (Sweden), MathCore (Sweden), Modelon (Sweden), Pöry (Finland), Qtronic (Germany), SICS (Sweden), Scania (Sweden), Semantum (Finland), Sherpa Engineering (France), Siemens (Germany and Sweden), Simpack (Germany), SKF (Sweden), Supmeca (France), Triphase (Belgium), University of Calabria (Italy), VTT (Finland), Vattenfall (Sweden), Wapice (Finland).

[0] http://www.s3pm.cominlabs.ueb.eu/

Abstract: Modelling and simulation are efficient and widely used tools for system design. But they are seldom used for systems operation. However, most functionalities for system design are beneficial for system operation, provided that they are enhanced to deal with real operating situations. Through open standards the benefits of sharing compatible information and data become obvious: improved cooperation between the design and the operation communities, easier adaptation of operation procedures wrt. design evolutions. Open standards also foster general purpose technology. The objective of the ITEA 2 MODRIO project is to extend modelling and simulation tools based on open standards from system design to system operation.

9. Dissemination

9.1. Promoting Scientific Activities

9.1.1. Scientific Events Selection

9.1.1.1. Member of the Conference Program Committees

Benoî Caillaud has served on the program committee of ACSD 2016 (http://acsd2016.mat.umk.pl), a conference on the applications of concurrency in system design. He is a member of the steering committe of ACSD since 2006.

9.1.1.2. Reviewer

Benoît Caillaud has reviewed papers submitted to the ACSD 2016 and ACC 2016 conferences.

Khalil Ghorbal reviewed two regular research papers for the Hybrid Systems: Computation and Control Conference.

Khalil Ghorbal reviewed two journal papers for the IEEE Transactions on Automatic Control.

Khalil Ghorbal reviewed a journal paper for the Computer Journal (Oxford Journals, Science and Mathematics).

Khalil Ghorbal reviewed a journal paper for the Information and Computation journal (Elsevier).

9.1.2. Invited Talks

Benoît Caillaud has given an invited talk on *Time Domains in Hybrid Systems Modeling* at the SHARC 2016 workshop and ALROB meetingthat took place in Brest in June 2016 (http://lab-sticc.univ-brest.fr/~goulven/sharc2016/program/index.html).

In May 13, 2016, Khalil Ghorbal gave an invited talk about the invariant generation for polynomial ordinary differential equations during the Effective Algebraic Geometry Seminar, IRMAR, Rennes, France.

In May 23, 2016, Ayman Aljarbouh presented a talk at the Embassy of Sweden in Tokyo for the first Japanese Modelica Conference (MODELICA2016), May 23-24, 2016, Tokyo, JAPAN.

In July 2016, Ayman Aljarbouh presented a poster during the French-American Doctoral Exchange Seminar (FADEx) 2016: Systèmes Cyber-Physiques, July 04-08, 2016, Grenoble, FRANCE.

In November, 5-12, Albert Benveniste was invited at the Systems Research Center, a center of excellence of the University of Maryland at College Park, USA.

9.1.3. Research Administration

Benoît Caillaud is head of the *Languages and Software Engineering* department of IRISA (http://www.irisa.fr/en/departments/d4-language-and-software-engineering). He has been in charge of presenting the departement during the evaluation seminar of IRISA by HCERES in January 2016.

9.2. Teaching - Supervision - Juries

9.2.1. Teaching

Master : Benoît Caillaud is teaching with Marc Pouzet a first year master degree course on *hybrid systems modeling*. The course is open to the students registered to the computer science research and innovation curriculum of the university of Rennes 1 and ENS Rennes, France.

Master : Khalil Ghorbal was "Chargé de TD" (20h Eq TD) for the "Analyse et Conception Formelles" module open for students registered to the computer science master degree of the university of Rennes 1 and ENS Rennes, France.

9.2.2. Supervision

PhD in progress : Ayman Aljarbouh, *Accelerated Simulation of Hybrid Systems*, started january 2014, supervised by Benoît Caillaud. Ayman Aljarbouh is expected to defend his PhD in MArch 2017.

9.2.3. Juries

Khalil Ghorbal was reviewer in the PhD defense committee of Sameh Mohamed, "Une Méthode Topologique pour la Recherche d'Ensembles Invariants de Systèmes Continus et á Commutation", defended in October 17th, 2016, Univ. Paris Saclay (ENS Cachan).

10. Bibliography

Major publications by the team in recent years

[1] A. Benveniste, T. Bourke, B. Caillaud, M. Pouzet.*Non-standard semantics of hybrid systems modelers*, in "Journal of Computer and System Sciences", 2012, vol. 78, n° 3, p. 877-910, This work was supported by the SYNCHRONICS large scale initiative of Inria [*DOI :* 10.1016/j.jcss.2011.08.009], http://hal.inria.fr/hal-00766726.

[2] A. Benveniste, B. Caillaud, D. Nickovic, R. Passerone, J.-B. Raclet, P. Reinkemeier, A. L. Sangiovanni-Vincentelli, W. Damm, T. A. Henzinger, K. G. Larsen.*Contracts for System Design*, Inria, November 2012, n° RR-8147, 65, http://hal.inria.fr/hal-00757488.

[3] J.-B. Raclet, E. Badouel, A. Benveniste, B. Caillaud, A. Legay, R. Passerone.*A Modal Interface Theory for Component-based Design*, in "Fundamenta Informaticae", 2011, vol. 108, n° 1-2, p. 119-149, http://dx.doi.org/10.3233/FI-2011-416.

Publications of the year

Articles in International Peer-Reviewed Journal

[4] G. Baudart, A. Benveniste, T. Bourke.*Loosely Time-Triggered Architectures*, in "ACM Transactions on Embedded Computing Systems (TECS)", August 2016, vol. 15, Article 71 [*DOI :* 10.1145/2932189], https://hal.inria.fr/hal-01408224.

[5] G. Claude, V. Gouranton, B. Caillaud, B. Gibaud, B. Arnaldi, P. Jannin.*Synthesis and Simulation of Surgical Process Models*, in "Studies in Health Technology and Informatics", 2016, vol. 220, p. 63–70 [*DOI :* 10.3233/978-1-61499-625-5-63], https://hal.archives-ouvertes.fr/hal-01300990.

[6] J.-B. JEANNIN, K. GHORBAL, Y. KOUSKOULAS, A. SCHMIDT, R. GARDNER, S. MITSCH, A. PLATZER.*A Formally Verified Hybrid System for Safe Advisories in the Next-Generation Airborne Collision Avoidance System*, in "International Journal on Software Tools for Technology Transfer", October 2016 [*DOI :* 10.1007/S10009-016-0434-1], https://hal.archives-ouvertes.fr/hal-01232365.

Invited Conferences

[7] A. ALJARBOUH, B. CAILLAUD.*Chattering-Free Simulation of Hybrid Dynamical Systems with the Functional Mock-Up Interface 2.0*, in "The First Japanese Modelica Conferences", Tokyo, Japan, Linköping University Electronic Press, Linköpings universitet, May 2016, vol. 124, n° 013, p. 95-105 [*DOI :* 10.3384/ECP1612495], https://hal.archives-ouvertes.fr/hal-01247008.

[8] A. ALJARBOUH, A. DURACZ, Y. ZENG, B. CAILLAUD, W. TAHA.*Chattering-Free Simulation for Hybrid Dynamical Systems: Semantics and Prototype Implementation*, in "2016 IEEE International Conference on Computational Science and Engineering, IEEE International Conference on Embedded and Ubiquitous Computing, and International Symposium on Distributed Computing and Applications to Business, Engineering and Science", Paris, France, Proceedings of the 19th IEEE International Conference on Computational Science and Engineering, the 14th IEEE International Conference on Embedded and Ubiquitous Computing, the 15th International Symposium on Distributed Computing and Applications to Business, Engineering and Science, IEEE Computer Society, August 2016 [*DOI :* 10.1109/CSE-EUC-DCABES.2016.217], https://hal.archives-ouvertes.fr/hal-01365875.

International Conferences with Proceedings

[9] G. CLAUDE, V. GOURANTON, B. CAILLAUD, B. GIBAUD, P. JANNIN, B. ARNALDI.*From Observations to Collaborative Simulation: Application to Surgical Training*, in "ICAT-EGVE 2016 - International Conference on Artificial Reality and Telexistence, Eurographics Symposium on Virtual Environments", Little Rock, Arkansas, United States, December 2016, https://hal.archives-ouvertes.fr/hal-01391776.

[10] A. T. SOGOKON, K. GHORBAL, T. T. JOHNSON.*Decoupling Abstractions of Non-linear Ordinary Differential Equations*, in "Formal Methods", Limassol, Cyprus, FM 2016: Formal Methods, November 2016, vol. Lecture Notes in Computer Science, n° 9995, p. 628-644 [*DOI :* 10.1007/978-3-319-48989-6_38], https://hal.archives-ouvertes.fr/hal-01374899.

Research Reports

[11] A. BENVENISTE, B. CAILLAUD, M. POUZET, H. ELMQVIST, M. OTTER.*Structural Analysis of Multi-Mode DAE Systems*, Inria, July 2016, n° RR-8933, 32, https://hal.inria.fr/hal-01343967.

References in notes

[12] N. J. CUTLAND (editor). *Nonstandard analysis and its applications*, Cambridge Univ. Press, 1988.

[13] *IEEE Standard VHDL Analog and Mixed-Signal Extensions, Std 1076.1-1999*, 1999, http://dx.doi.org/10.1109/IEEESTD.1999.90578.

[14] A. ANTONIK, M. HUTH, K. G. LARSEN, U. NYMAN, A. WASOWSKI.*20 Years of Modal and Mixed Specifications*, in "Bulletin of European Association of Theoretical Computer Science", 2008, vol. 1, n° 94.

[15] C. BAIER, J.-P. KATOEN.*Principles of Model Checking*, MIT Press, Cambridge, 2008.

[16] A. BEN KHALED, M. E. M. BEN GAÏD, N. PERNET, D. SIMON.*Fast multi-core co-simulation of Cyber-Physical Systems : application to internal combustion engines*, in "Simulation Modelling Practice and Theory", September 2014, vol. 47, n^o September, p. 79-91 [*DOI :* 10.1016/J.SIMPAT.2014.05.002], https://hal-ifp.archives-ouvertes.fr/hal-01018348.

[17] A. BENVENISTE, T. BOURKE, B. CAILLAUD, B. PAGANO, M. POUZET.*A Type-Based Analysis of Causality Loops In Hybrid Systems Modelers*, December 2013, Deliverable D3.1_1 v 1.0 of the Sys2soft collaborative project "Physics Aware Software", https://hal.inria.fr/hal-00938866.

[18] A. BENVENISTE, T. BOURKE, B. CAILLAUD, M. POUZET.*Semantics of multi-mode DAE systems*, August 2013, Deliverable D.4.1.1 of the ITEA2 Modrio collaborative project, https://hal.inria.fr/hal-00938891.

[19] A. BENVENISTE, B. CAILLAUD, H. ELMQVIST, K. GHORBAL, M. OTTER, M. POUZET.*Structural Analysis of Multi-Mode DAE Systems*, in "Proc. of the 20th ACM International Conference on Hybrid Systems: Computation and Control, HSCC'17", Pittsburgh, PA, USA, April 2017, to appear.

[20] A. BENVENISTE, B. CAILLAUD, A. FERRARI, L. MANGERUCA, R. PASSERONE, C. SOFRONIS.*Multiple Viewpoint Contract-Based Specification and Design*, in "Proceedings of the Software Technology Concertation on Formal Methods for Components and Objects (FMCO'07)", Amsterdam, The Netherlands, Revised Lectures, Lecture Notes in Computer Science, Springer, October 2008, vol. 5382.

[21] A. BENVENISTE, B. CAILLAUD, B. PAGANO, M. POUZET.*A type-based analysis of causality loops in hybrid modelers*, in "HSCC '14: International Conference on Hybrid Systems: Computation and Control", Berlin, Germany, Proceedings of the 17th international conference on Hybrid systems: computation and control (HSCC '14), ACM Press, April 2014, 13 [*DOI :* 10.1145/2562059.2562125], https://hal.inria.fr/hal-01093388.

[22] N. BERTRAND, A. LEGAY, S. PINCHINAT, J.-B. RACLET.*A Compositional Approach on Modal Specifications for Timed Systems*, in "11th International Conference on Formal Engineering Methods (ICFEM'09)", Rio de Janeiro, Brazil, LNCS, Springer, December 2009, vol. 5885, p. 679-697, http://hal.inria.fr/inria-00424356/en.

[23] N. BERTRAND, A. LEGAY, S. PINCHINAT, J.-B. RACLET.*Modal event-clock specifications for timed component-based design*, in "Science of Computer Programming", 2011, http://dx.doi.org/10.1016/j.scico.2011.01.007.

[24] N. BERTRAND, S. PINCHINAT, J.-B. RACLET.*Refinement and Consistency of Timed Modal Specifications*, in "3rd International Conference on Language and Automata Theory and Applications (LATA'09)", Tarragona, Spain, LNCS, Springer, April 2009, vol. 5457, p. 152-163 [*DOI :* 10.1007/978-3-642-00982-2_13], http://hal.inria.fr/inria-00424283/en.

[25] P. BHADURI, I. STIERAND.*A proposal for real-time interfaces in SPEEDS*, in "Design, Automation and Test in Europe (DATE'10)", IEEE, 2010, p. 441-446.

[26] G. BISWAS, A. BREGON, X. KOUTSOUKOS, B. PULIDO.*Analytic Redundancy, Possible Conflicts, and TCG-based Fault Signature Diagnosis applied to Nonlinear Dynamic Systems*, in "IFAC Proceedings Vol-

umes", 2009, vol. 42, n^{o} 8, p. 1486 - 1491 [*DOI :* 10.3182/20090630-4-ES-2003.00242], http://www.sciencedirect.com/science/article/pii/S1474667016359857.

[27] S. BLIUDZE.*Un cadre formel pour l'étude des systèmes industriels complexes: un exemple basé sur l'infrastructure de l'UMTS*, Ecole Polytechnique, 2006.

[28] S. BLIUDZE, D. KROB.*Modelling of Complex Systems: Systems as Dataflow Machines*, in "Fundam. Inform.", 2009, vol. 91, n^{o} 2, p. 251–274.

[29] G. BOUDOL, K. G. LARSEN.*Graphical Versus Logical Specifications*, in "Theor. Comput. Sci.", 1992, vol. 106, n^{o} 1, p. 3-20.

[30] D. BROMAN, L. GREENBERG, E. A. LEE, M. MASIN, S. TRIPAKIS, M. WETTER.*Requirements for hybrid cosimulation standards*, in "Proceedings of the 18th International Conference on Hybrid Systems: Computation and Control, HSCC'15, Seattle, WA, USA, April 14-16, 2015", A. GIRARD, S. SANKARANARAYANAN (editors), ACM, 2015, p. 179–188, http://doi.acm.org/10.1145/2728606.2728629.

[31] B. CAILLAUD.*Surgical Process Mining with Test and Flip Net Synthesis*, in "Application of Region Theory (ART)", Barcelona, Spain, R. BERGENTHUM, J. CARMONA (editors), July 2013, p. 43-54, http://hal.inria.fr/hal-00872284.

[32] B. CAILLAUD, B. DELAHAYE, K. G. LARSEN, A. LEGAY, M. L. PEDERSEN, A. WASOWSKI.*Compositional design methodology with constraint Markov chains*, in "QEST 2010", Williamsburg, Virginia, United States, September 2010 [*DOI :* 10.1109/QEST.2010.23], http://hal.inria.fr/inria-00591578/en.

[33] B. CAILLAUD, B. DELAHAYE, K. G. LARSEN, A. LEGAY, M. L. PEDERSEN, A. WASOWSKI.*Constraint Markov Chains*, in "Theoretical Computer Science", May 2011, vol. 412, n^{o} 34, p. 4373-4404 [*DOI :* 10.1016/J.TCS.2011.05.010], http://hal.inria.fr/hal-00654003/en.

[34] A. CHAKRABARTI.*A Framework for Compositional Design and Analysis of Systems*, EECS Department, University of California, Berkeley, Dec 2007, http://www.eecs.berkeley.edu/Pubs/TechRpts/2007/EECS-2007-174.html.

[35] A. CHAKRABARTI, L. DE ALFARO, T. A. HENZINGER, M. STOELINGA.*Resource Interfaces*, in "EMSOFT", R. ALUR, I. LEE (editors), Lecture Notes in Computer Science, Springer, 2003, vol. 2855, p. 117-133.

[36] E. Y. CHANG, Z. MANNA, A. PNUELI.*Characterization of Temporal Property Classes*, in "ICALP", W. KUICH (editor), Lecture Notes in Computer Science, Springer, 1992, vol. 623, p. 474-486.

[37] E. CLARKE, O. GRUMBERG, D. PELED.*Model Checking*, MIT Press, 1999.

[38] W. DAMM, E. THADEN, I. STIERAND, T. PEIKENKAMP, H. HUNGAR.*Using Contract-Based Component Specifications for Virtual Integration and Architecture Design*, in "Proceedings of the 2011 Design, Automation and Test in Europe (DATE'11)", March 2011.

[39] A. DAVID, K. G. LARSEN, A. LEGAY, U. NYMAN, A. WASOWSKI.*ECDAR: An Environment for Compositional Design and Analysis of Real Time Systems*, in "Automated Technology for Verification and Analysis -

8th International Symposium, ATVA 2010, Singapore, September 21-24, 2010. Proceedings", 2010, p. 365-370.

[40] A. DAVID, K. G. LARSEN, A. LEGAY, U. NYMAN, A. WASOWSKI.*Timed I/O automata: a complete specification theory for real-time systems*, in "Proceedings of the 13th ACM International Conference on Hybrid Systems: Computation and Control, HSCC 2010, Stockholm, Sweden, April 12-15, 2010", 2010, p. 91-100.

[41] B. DELAHAYE, J.-P. KATOEN, K. G. LARSEN, A. LEGAY, M. L. PEDERSEN, F. SHER, A. WASOWSKI.*Abstract Probabilistic Automata*, in "VMCAI", R. JHALA, D. A. SCHMIDT (editors), Lecture Notes in Computer Science, Springer, 2011, vol. 6538, p. 324-339.

[42] F. DIENER, G. REEB.*Analyse non standard*, Hermann, 1989.

[43] D. L. DILL.*Trace Theory for Automatic Hierarchical Verification of Speed-Independent Circuits*, ACM Distinguished Dissertations, MIT Press, 1989.

[44] H. ELMQVIST, S. E. MATSSON, H. OLSSON.*Parallel Model Execution on Many Cores*, in "Proceedings of the 10th International Modelica Conference; March 10-12; 2014; Lund; Sweden", Linköping University Electronic Press; Linköpings universitet, 2014, p. 363-370.

[45] V. GALTIER, S. VIALLE, C. DAD, J.-P. TAVELLA, J.-P. LAM-YEE-MUI, G. PLESSIS.*FMI-Based Distributed Multi-Simulation with DACCOSIM*, in "Symposium on Theory of Modeling and Simulation - TMS'15", Alexandria, VA, United States, April 2015, p. 804-811, https://hal-supelec.archives-ouvertes.fr/hal-01155707.

[46] Y. IWASAKI, A. FARQUHAR, V. SARASWAT, D. BOBROW, V. GUPTA.*Modeling Time in Hybrid Systems: How Fast Is "Instantaneous"?*, in "IJCAI", 1995, p. 1773–1781.

[47] L. LAMPORT.*Proving the Correctness of Multiprocess Programs*, in "IEEE Trans. Software Eng.", 1977, vol. 3, n^o 2, p. 125-143.

[48] K. G. LARSEN, U. NYMAN, A. WASOWSKI.*On Modal Refinement and Consistency*, in "Proc. of the 18th International Conference on Concurrency Theory (CONCUR'07)", Springer, 2007, p. 105–119.

[49] K. G. LARSEN, B. THOMSEN.*A Modal Process Logic*, in "Proceedings of the Third Annual Symposium on Logic in Computer Science (LICS'88)", IEEE, 1988, p. 203-210.

[50] T. LINDSTRØM.*An Invitation to Nonstandard Analysis*, in "Nonstandard Analysis and its Applications", N. J. CUTLAND (editor), Cambridge Univ. Press, 1988, p. 1–105.

[51] N. A. LYNCH.*Input/Output Automata: Basic, Timed, Hybrid, Probabilistic, Dynamic, ...*, in "CONCUR", R. M. AMADIO, D. LUGIEZ (editors), Lecture Notes in Computer Science, Springer, 2003, vol. 2761, p. 187-188.

[52] N. A. LYNCH, E. W. STARK.*A Proof of the Kahn Principle for Input/Output Automata*, in "Inf. Comput.", 1989, vol. 82, n^o 1, p. 81-92.

[53] Z. MANNA, A. PNUELI.*Temporal verification of reactive systems: Safety*, Springer, 1995.

[54] B. MEYER.*Applying "Design by Contract"*, in "Computer", October 1992, vol. 25, n^o 10, p. 40–51, http://dx.doi.org/10.1109/2.161279.

[55] P. NUZZO, A. L. SANGIOVANNI-VINCENTELLI, X. SUN, A. PUGGELLI.*Methodology for the Design of Analog Integrated Interfaces Using Contracts*, in "IEEE Sensors Journal", Dec. 2012, vol. 12, n^o 12, p. 3329–3345.

[56] A. ROBINSON.*Non-Standard Analysis*, Princeton Landmarks in Mathematics, 1996, ISBN 0-691-04490-2.

[57] E. SIKORA, B. TENBERGEN, K. POHL.*Industry needs and research directions in requirements engineering for embedded systems*, in "Requirements Engineering", 2012, vol. 17, p. 57–78, http://link.springer.com/article/10.1007/s00766-011-0144-x.

[58] L. DE ALFARO.*Game Models for Open Systems*, in "Verification: Theory and Practice", Lecture Notes in Computer Science, Springer, 2003, vol. 2772, p. 269-289.

[59] L. DE ALFARO, T. A. HENZINGER.*Interface automata*, in "Proc. of the 9th ACM SIGSOFT International Symposium on Foundations of Software Engineering (FSE'01)", ACM Press, 2001, p. 109–120.

[60] L. DE ALFARO, T. A. HENZINGER.*Interface-based design*, in "In Engineering Theories of Software Intensive Systems, proceedings of the Marktoberdorf Summer School", Kluwer, 2004.

[61] L. DE ALFARO, T. A. HENZINGER, M. STOELINGA.*Timed Interfaces*, in "Proc. of the 2nd International Workshop on Embedded Software (EMSOFT'02)", Lecture Notes in Computer Science, Springer, 2002, vol. 2491, p. 108–122.

Project-Team I4S

Statistical Inference for Structural Health Monitoring

IN PARTNERSHIP WITH:
Institut français des sciences et technologies des transports, de l'aménagement et des réseaux - IFSTTAR

RESEARCH CENTER
Rennes - Bretagne-Atlantique

THEME
Optimization and control of dynamic systems

Table of contents

Project-Team I4S

Creation of the Team: 2009 April 01, updated into Project-Team: 2013 January 01

Keywords:

Computer Science and Digital Science:

6.1.5. - Multiphysics modeling
6.2.1. - Numerical analysis of PDE and ODE
6.2.4. - Statistical methods
6.2.5. - Numerical Linear Algebra
6.2.6. - Optimization
6.3.1. - Inverse problems
6.3.3. - Data processing
6.3.4. - Model reduction
6.3.5. - Uncertainty Quantification
6.4.3. - Observability and Controlability

Other Research Topics and Application Domains:

3.1. - Sustainable development
3.2. - Climate and meteorology
3.3.1. - Earth and subsoil
4.3.2. - Hydro-energy
4.3.3. - Wind energy
4.3.4. - Solar Energy
5.1. - Factory of the future
5.2. - Design and manufacturing
5.9. - Industrial maintenance
6.5. - Information systems
7.2.2. - Smart road
8.1. - Smart building/home
8.1.1. - Energy for smart buildings
8.1.2. - Sensor networks for smart buildings
8.2. - Connected city

1. Members

Research Scientists

Laurent Mevel [Team leader, Inria, Senior Researcher, HDR]
Michael Doehler [Inria, Researcher]
Jean Dumoulin [IFSTTAR, Researcher]
Dominique Siegert [IFSTTAR, Researcher, until Dec 2016]
Xavier Chapeleau [IFSTTAR, Researcher, since January 2016]
Qinghua Zhang [Inria, Senior Researcher, HDR]

Technical Staff

Antoine Crinière [Inria]

Ivan Guéguen [IFSTTAR]
Vincent Le Cam [IFSTTAR, Senior Engineer]
Mathieu Le Pen [IFSTTAR]

PhD Students

Nassif Berrabah [EDF, granted by CIFRE]
Md Delwar Hossain Bhuyan [Inria, granted by ARED Bretagne]
Nicolas Le Touz [IFSTTAR]
Thibaud Toullier [Inria, from Oct 2016, granted by ARED Bretagne]
Eva Viefhues [BAM, from Sep 2016]
Saeid Allahdadian [UBC, from Sep 2016]
Antoine Bassil [IFSTTAR, from Nov 2016]

Post-Doctoral Fellows

Jordan Brouns [IFSTTAR, until Mar 2016]
Guillaume Gautier [CEA]
Subhamoy Sen [Inria, from Nov 2016]

Other

Youssef Elkamali [Inria,Student,from Apr 2016 until Jul 2016]

2. Overall Objectives

2.1. Overall Objectives

monitoring, system identification, on-line identification and detection algorithms, statistical hypotheses testing, reflectometry, infrared thermography, non destructive testing, sensors fusion, optimal sensors placement, vibration-based structural analysis and damage detection and localization, aeronautics, civil engineering

2.1.1. In Summary

The objective of this team is the development of Structural Health Monitoring techniques by intrinsic coupling of statistics and thermo-aeroelastic mixing modeling for the development of robust and autonomous structural health monitoring solutions of mechanical structures. The emphasis of the team is the handling of very large systems such as the recent wind energy converters currently being installed in Europe, building on the expertise acquired by the team on bridges as an example of civil engineering structure, and for aircrafts and helicopters in the context of aero elastic instability monitoring. The necessity of system identification and damage detection systems robust to environmental variations and being designed to handle a very large model dimension motivates us. As examples, the explosion in the installed number of sensors and the robustness to temperature variation will be the main focus of the team. This implies new statistical and numerical technologies as well as improvements on the modeling of the underlying physical models. Many techniques and methods originate from the mechanical community and thus exhibit a very deep understanding of the underlying physics and mechanical behavior of the structure. On the other side, system identification techniques developed within the control community are more related to data modeling and take into account the underlying random nature of measurement noise. Bringing these two communities together is the objective of this joint team between Inria and IFSTTAR. It will results hopefully in methods numerically robust, statistically efficient and also mixing modeling of both the uncertainties related to the data and the associated complex physical models related to the laws of physics and finite element models.

Damage detection in civil structures has been a main focus over the last decade. Still, those techniques need to be matured to be operable and installed on structures in operation, and thus be robust to environmental nuisances. Then, damage localization, quantification and prognosis should be in that order addressed by the team. To be precise and efficient, it requires correct mixing between signal processing, statistical analysis, Finite Elements Models (FEM) updating and a yet to be available precise modeling of the environmental effects such as temperature through 3D field reconstruction.

Theoretical and practical questions are more and more complex. For example, in civil engineering, from handling hundreds of sensors automatically during some long period of time to localize and quantify damage with or without numerical models. Very large heavily instrumented structures are yet to come and they will ask for a paradigm in how we treat them from a renewed point of view. As the structures become large and complex, also the thermal and aeroelastic (among others) models become complex. Bridges and aircrafts are the main focus of our research. Opening our expertise on new applications topics such as helicopters and wind energy converters is also part of our priorities.

2.1.1.1. Objectives

The main objectives of the team are first to pursue current algorithmic research activities, in order to accommodate still-to-be-developed complex physical models. More precisely, we want successively

- To develop statistical algorithms robust to noise and variation in the environment
- To handle transient and highly varying systems under operational conditions
- To consider the impact of uncertainties on the current available identification algorithms and develop efficient, robust and fast implementation of such quantities
- To consider relevant non trivial thermal models for usage in rejection based structural health monitoring and more generally to mix numerical model, physical modeling and data
- To develop theoretical and software tools for monitoring and localization of damages on civil structures or instability for aircrafts
- To explore new paradigms for handling very large and complex structures heavily instrumented (distributed computing)
- To study the characteristics of the monitored mechanic structures in terms of electromagnetic propagation, in order to develop monitoring methods based on electrical instrumentations.
- To consider society concerns (damage quantification and remaining life prognosis)

2.1.2. Introduction to physics driven dynamical models in the context of civil engineering elastic structures

The design and maintenance of flexible structures subject to noise and vibrations is an important topic in civil and mechanical engineering. It is an important component of comfort (cars and buildings) and contributes significantly to the safety related aspects of design and maintenance (aircrafts, aerospace vehicles and payloads, long-span bridges, high-rise towers...). Requirements from these application areas are numerous and demanding.

Detailed physical models derived from first principles are developed as part of system design. These models involve the dynamics of vibrations, sometimes complemented by other physical aspects (fluid-structure interaction, aerodynamics, thermodynamics).

Laboratory and in-operation tests are performed on mock-up or real structures, in order to get so-called modal models, ie to extract the modes and damping factors (these correspond to system poles), the mode shapes (corresponding eigenvectors), and loads. These results are used for updating the design model for a better fit to data, and sometimes for certification purposes (e.g. in flight domain opening for new aircrafts, reception for large bridges).

The monitoring of structures is an important activity for the system maintenance and health assessment. This is particularly important for civil structures. Damaged structures would typically exhibit often very small changes in their stiffness due to the occurrence of cracks, loss of prestressing or post tensioning, chemical reactions, evolution of the bearing behavior and most importantly scour. A key difficulty is that such system characteristics are also sensitive to environmental conditions, such as temperature effects (for civil structures), or external loads (for aircrafts). In fact these environmental effects usually dominate the effect of damage. This is why, for very critical structures such as aircrafts, detailed active inspection of the structures is performed as part of the maintenance. Of course, whenever modal information is used to localize a damage, the localization of a damage should be expressed in terms of the physical model, not in terms of the modal model used in system identification. Consequently, the following elements are encountered and must be jointly dealt with when addressing these applications: design models from the system physics, modal models used in structural identification, and, of course, data from sensors. Corresponding characteristics are given now: Design models are Finite Element models, sometimes with tens or hundreds of thousands elements, depending on professional habits which may vary from one sector to another. These models are linear if only small vibrations are considered; still, these models can be large if medium-frequency spectrum of the load is significant. In addition, nonlinearities enter as soon as large vibrations or other physical effects (aerodynamics, thermodynamics, ..) are considered. Moreover stress-strain paths and therefore the response (and load) history comes into play.

Sensors can range from a handful of accelerometers or strain gauges, to thousands of them, if NEMS (Nano Electro Mechanical Structures), MEMS (Microelectromechanical systems) or optical fiber sensors are used. Moreover, the sensor output can be a two-dimensional matrix if electro magnet (IR (infrared), SAR, shearography ...) or other imaging technologies are used.

2.1.2.1. *Multi-fold thermal effects*

The temperature constitutes an often dominant load because it can generate a deflection as important as that due to the self-weight of a bridge. In addition, it sometimes provokes abrupt slips of bridge spans on their bearing devices, which can generate significant transient stresses as well as a permanent deformation, thus contributing to fatigue.

But it is also well-known that the dynamic behavior of structures under monitoring can vary under the influence of several factors, including the temperature variations, because they modify the stiffness and thus the modes of vibration. As a matter of fact, depending on the boundary conditions of the structure, possibly uniform thermal variations can cause very important variations of the spectrum of the structure, up to 10%, because in particular of additional prestressing, not forgetting pre strain, but also because of the temperature dependence of the characteristics of materials. As an example, the stiffness of elastomeric bearing devices vary considerably in the range of extreme temperatures in some countries. Moreover, eigenfrequencies and modal shapes do not depend monotonically with temperature. Abrupt dynamical behavior may show up due to a change of boundary conditions e.g. due to limited expansion or frost bearing devices. The temperature can actually modify the number of contact points between the piles and the main span of the bridge. Thus the environmental effects can be several orders of magnitude more important than the effect of true structural damages. It will be noted that certain direct methods aiming at detecting local curvature variations stumble on the dominating impact of the thermal gradients. In the same way, the robustness and effectiveness of model-based structural control would suffer from any unidentified modification of the vibratory behavior of the structure of interest. Consequently, it is mandatory to cure dynamic sensor outputs from thermal effects before signal processing can help with a diagnostics on the structure itself, otherwise the possibility of reliable ambient vibration monitoring of civil structures remains questionable. Despite the paramount interest this question deserves, thermal elimination still appears to challenge the SHM community.

2.1.2.2. *Toward a multidisciplinary approach*

Unlike previously mentioned blind approaches, successful endeavours to eliminate the temperature from subspace-based damage detection algorithms prove the relevance of relying on predictive thermo-mechanical models yielding the prestress state and associated strains due to temperature variations. As part of the CONSTRUCTIF project supported by the Action Concertée Incitative Sécurité Informatique of the French Ministry for Education and Research, very encouraging results in this direction were obtained and published. they were substantiated by laboratory experiments of academic type on a simple beam subjected to a known uniform temperature. Considering the international pressure toward reliable methods for thermal elimination, these preliminary results pave the ground to a new SHM paradigm. Moreover, for one-dimensional problems, it was shown that real time temperature identification based on optimal control theory is possible provided the norm of the reconstructed heat flux is properly chosen. Finally, thermo-mechanical models of vibrating thin structures subject to thermal prestress, prestrain, geometric imperfection and damping have been extensively revisited. This project led by Inria involved IFSTTAR where the experiments were carried out. The project was over in July 2006. Note that thermo-mechanics of bridge piles combined with an *ad hoc* estimation of thermal gradients becomes of interest to practicing engineers. Thus, I4S's approach should suit advanced professional practice. Finite element analysis is also used to predict stresses and displacements of large bridges in Hong-Kong bay .

Temperature rejection is the primary focus and obstacle for SHM projects I4S participates in civil engineering, like SIMS project in Canada, ISMS in Danemark or SIPRIS in France.

A recent collaboration between Inria and IFSTTAR has demonstrated the efficiency of reflectometry-based methods for health monitoring of some civil engineering structures, notably external post-tensioned cables. Based on a mathematical model of electromagnetic propagation in mechanical structures, the measurement of reflected and transmitted electromagnetic waves by the monitored structures allows to detect structural failures. The interaction of such methods with those based on mechanical and thermal measurements will reinforce the multidisciplinary approach developed in our team.

2.1.2.3. Models for monitoring under environmental changes - scientific background

We will be interested in studying linear stochastic systems, more precisely, assume at hand a sequence of observations Y_n measured during time,

$$\left\{ \begin{array}{rcl} X_{n+1} & = & AX_n + V_n \\ Y_n & = & HX_n + W_n \end{array} \right. \tag{4}$$

where V_n and W_n are zero mean random variables, A is the transition matrix of the system, H is the observation matrix between state and observation, and X_n the process describing the monitored system. X_n can be related to a physical process (for example, for a mechanical structure, the collection of displacements and velocities at different points). Different problems arise

1/ identify and characterize the structure of interest. It may be possible by matching a parametric model to the observed time series Y_n in order to minimize some given criterion, whose minimum will be the best approximation describing the system,

2/ decide if the measured data describe a system in a so called "reference" state (the term "reference" is used in the context of fault detection, where the reference is considered to be safe) and monitor its deviations with respect of its nominal reference state.

Both problems should be addressed differently if

1/ we consider that the allocated time to measurement is large enough, resulting in a sequence Y_n whose size tends to infinity, a requirement for obtaining statistical convergence results. It corresponds to the identification and monitoring of a dynamical system with slow variations. For example, this description is well suited to the long-term monitoring of civil structures, where records can be measured during relatively (to sampling rate) large periods of time (typically many minutes or hours).

2/ we are interested in systems, whose dynamic is fast with respect to the sampling rate, most often asking for reaction in terms of seconds. It is, for example, the case for mission critical applications such as in-flight control or real-time security and safety assessment. Both aeronautics and transport or utilities infrastructures are concerned. In this case, fast algorithms with sample-by-sample reaction are necessary.

The monitoring of mechanical structures can not be addressed without taking into account the close environment of the considered system and their interactions. Typically, monitored structures of interest do not reside in laboratory but are considered in operational conditions, undergoing temperature, wind and humidity variations, as well as traffic, water flows and other natural or man-made loads. Those variations do imply a variation of the eigenproperties of the monitored structure, variations to be separated from the damage/instability induced variations.

For example, in civil engineering, an essential problem for in operation health monitoring of civil structures is the variation of the environment itself. Unlike laboratory experiments, civil structure modal properties change during time as temperature and humidity vary. Traffic and comparable transient events also influence the structures. Thus, structural modal properties are modified by slow low variations, as well as fast transient non stationarities. From a damage detection point of view, the former has to be detected, whereas the latter has to be neglected and not perturb the detection. Of course, from a structural health monitoring point of view the knowledge of the true load is in itself of paramount importance.

In this context, the considered perturbations will be of two kinds, either

1/ the influence of the temperature on civil structures, such as bridges or wind energy converters : as we will notice, those induced variations can be modeled by a additive component on the system stiffness matrix depending on the current temperature, as

$$K = K_{struct} + K_T \ .$$

We will then have to monitor the variations in K_{struct} independently of the variations in K_T, based on some measurements generated from a system, whose stiffness matrix is K.

2/ the influence of the aeroelastic forces on aeronautical structures such as aircrafts or rockets and on flexible civil structures such as long-span bridges : we will see as well that this influence implies a modification of the classical mechanical equation (2)

$$M\ddot{Z} + C\dot{Z} + KZ = V \tag{5}$$

where (M, C, K) are the mass, damping and stiffness matrices of the system and Z the associated vector of displacements measured on the monitored structure. In a first approximation, those quantities are related by (2). Assuming U is the velocity of the system, adding U dependent aeroelasticity terms, as in (3), introduces a coupling between U and (M, C, K).

$$M\ddot{Z} + C\dot{Z} + KZ = U^2 DZ + UE\dot{Z} + V \tag{6}$$

Most of the research at Inria for a decade has been devoted to the study of subspace methods and how they handle the problems described above.

Model (2) is characterized by the following property (we formulate it for the single sensor case, to simplify notations): Let $y_{-N} \cdots y_{+N}$ be the data set, where N is large, and let M, P sufficiently smaller than N for the following objects to make sense: 1/ define the row vectors $Y_k = (y_k \cdots y_{k-M}), |k| \leq P$; 2/ stack the Y_k on top of each other for $k = 0, 1, \cdots, P$ to get the data matrix $\mathcal{Y}_+$ and stack the column vectors Y_k^T for $k = 0, -1, \cdots, -P$ to get the data matrix $\mathcal{Y}_-$; 3/ the product $\mathcal{H} = \mathcal{Y}_+ \mathcal{Y}_-$ is a Hankel matrix. Then, matrix $\mathcal{H}$ on the one hand, and the observability matrix $\mathcal{O}(H, F)$ of system (2) on the other hand, possess almost identical left kernel spaces, asymptotically for M, N large. This property is the basis of subspace identification methods. Extracting $\mathcal{O}(H, F)$ using some Singular Value Decomposition from $\mathcal{H}$ then (H, F) from $\mathcal{O}(H, F)$ using a Least Square approach has been the foundation of the academic work on subspace methods for many years. The team focused on the numerical efficiency and consistency of those methods and their applicability on solving the problems above.

There are numerous ways to implement those methods. This approach has seen a wide acceptance in the industry and benefits from a large background in the automatic control literature. Up to now, there was a discrepancy between the a priori efficiency of the method and some not so efficient implementations of this algorithm. In practice, for the last ten years, stabilization diagrams have been used to handle the instability and the weakness with respect to noise, as well as the poor capability of those methods to determine model orders from data. Those methods implied some engineering expertise and heavy post processing to discriminate between models and noise. This complexity has led the mechanical community to adopt preferably frequency domain methods such as Polyreference LSCF. Our focus has been on improving the numerical stability of the subspace algorithms by studying how to compute the least square solution step in this algorithm. This yields to a very efficient noise free algorithm, which has provided a renewed acceptance in the mechanical engineering community for the subspace algorithms. Now we focus on improving speed and robustness of those algorithms.

Subspace methods can also be used to test whether a given data set conforms a model: just check whether this property holds, for a given pair {data, model}. Since equality holds only asymptotically, equality must be tested against some threshold ε; tuning ε relies on so-called *asymptotic local* approach for testing between close hypotheses on long data sets — this method was introduced by Le Cam in the 70s. By using the Jacobian between pair (H, F) and the modes and mode shapes, or the Finite Element Model parameters, one can localize and assess the damage.

In oder to discriminate between damage and temperature variations, we need to monitor the variations in K_{struct} while keeping blind to the variations in K_T in statistical terms, we must detect and diagnose changes in K_{struct} while rejecting nuisance parameter K_T. Several techniques were explored in the thesis of Houssein Nasser, from purely empirical approaches to (physical) model based approaches. Empirical approaches do work, but model based approaches are the most promising and a focus of our future researches. This approach requires a physical model of how temperature affects stiffness in various materials. This is why a large part of our future research is devoted to the modeling of such environmental effect.

This approach has been used also for flutter monitoring in Rafik Zouari's PhD thesis for handling the aeroelastic effect.

3. Research Program

3.1. Vibration analysis

In this section, the main features for the key monitoring issues, namely identification, detection, and diagnostics, are provided, and a particular instantiation relevant for vibration monitoring is described.

It should be stressed that the foundations for identification, detection, and diagnostics, are fairly general, if not generic. Handling high order linear dynamical systems, in connection with finite elements models, which call for using subspace-based methods, is specific to vibration-based SHM. Actually, one particular feature of model-based sensor information data processing as exercised in I4S, is the combined use of black-box or semi-physical models together with physical ones. Black-box and semi-physical models are, for example, eigenstructure parameterizations of linear MIMO systems, of interest for modal analysis and vibration-based SHM. Such models are intended to be identifiable. However, due to the large model orders that need to be considered, the issue of model order selection is really a challenge. Traditional advanced techniques from statistics such as the various forms of Akaike criteria (AIC, BIC, MDL, ...) do not work at all. This gives rise to new research activities specific to handling high order models.

Our approach to monitoring assumes that a model of the monitored system is available. This is a reasonable assumption, especially within the SHM areas. The main feature of our monitoring method is its intrinsic ability to the early warning of small deviations of a system with respect to a reference (safe) behavior under usual operating conditions, namely without any artificial excitation or other external action. Such a normal behavior is summarized in a reference parameter vector θ_0, for example a collection of modes and mode-shapes.

3.1.1. Identification

The behavior of the monitored continuous system is assumed to be described by a parametric model $\{\mathbf{P}_\theta \,,\, \theta \in \Theta\}$, where the distribution of the observations $(Z_0, ..., Z_N)$ is characterized by the parameter vector $\theta \in \Theta$.

For reasons closely related to the vibrations monitoring applications, we have been investigating subspace-based methods, for both the identification and the monitoring of the eigenstructure (λ, ϕ_λ) of the state transition matrix F of a linear dynamical state-space system :

$$\left\{ \begin{array}{rcl} X_{k+1} & = & F\; X_k + V_{k+1} \\ Y_k & = & H\; X_k + W_k \end{array} \right. , \tag{7}$$

namely the $(\lambda, \varphi_\lambda)$ defined by :

$$\det\ (F - \lambda\ I) = 0, \quad (F - \lambda\ I)\ \phi_\lambda = 0, \quad \varphi_\lambda \stackrel{\Delta}{=} H\ \phi_\lambda \tag{8}$$

The (canonical) parameter vector in that case is :

$$\theta \stackrel{\Delta}{=} \begin{pmatrix} \Lambda \\ \mathrm{vec}\Phi \end{pmatrix} \tag{9}$$

where Λ is the vector whose elements are the eigenvalues λ, Φ is the matrix whose columns are the φ_λ's, and vec is the column stacking operator.

Subspace-based methods is the generic name for linear systems identification algorithms based on either time domain measurements or output covariance matrices, in which different subspaces of Gaussian random vectors play a key role [54].

Let $R_i \stackrel{\Delta}{=} \mathbf{E}\left(Y_k\ Y_{k-i}^T\right)$ and:

$$\mathcal{H}_{p+1,q} \stackrel{\Delta}{=} \begin{pmatrix} R_1 & R_2 & \vdots & R_q \\ R_2 & R_3 & \vdots & R_{q+1} \\ \vdots & \vdots & \vdots & \vdots \\ R_{p+1} & R_{p+2} & \vdots & R_{p+q} \end{pmatrix} \stackrel{\Delta}{=} \mathrm{Hank}\,(R_i) \tag{10}$$

be the output covariance and Hankel matrices, respectively; and: $G \stackrel{\Delta}{=} \mathbf{E}\left(X_k Y_{k-1}^T\right)$. Direct computations of the R_i's from the equations (4) lead to the well known key factorizations :

$$\begin{array}{rcl} R_i & = & HF^{i-1}G \\ \mathcal{H}_{p+1,q} & = & \mathcal{O}_{p+1}(H,F)\ \mathcal{C}_q(F,G) \end{array} \tag{11}$$

where:

$$\mathcal{O}_{p+1}(H,F) \stackrel{\Delta}{=} \begin{pmatrix} H \\ HF \\ \vdots \\ HF^p \end{pmatrix} \quad \text{and} \quad \mathcal{C}_q(F,G) \stackrel{\Delta}{=} (G\ FG\ \cdots\ F^{q-1}G) \tag{12}$$

are the observability and controllability matrices, respectively. The observation matrix H is then found in the first block-row of the observability matrix $\mathcal{O}$. The state-transition matrix F is obtained from the shift invariance property of $\mathcal{O}$. The eigenstructure (λ, ϕ_λ) then results from (5).

Since the actual model order is generally not known, this procedure is run with increasing model orders.

3.1.2. Detection

Our approach to on-board detection is based on the so-called asymptotic statistical local approach. It is worth noticing that these investigations of ours have been initially motivated by a vibration monitoring application example. It should also be stressed that, as opposite to many monitoring approaches, our method does not require repeated identification for each newly collected data sample.

For achieving the early detection of small deviations with respect to the normal behavior, our approach generates, on the basis of the reference parameter vector θ_0 and a new data record, indicators which automatically perform :

- The early detection of a slight mismatch between the model and the data;
- A preliminary diagnostics and localization of the deviation(s);
- The tradeoff between the magnitude of the detected changes and the uncertainty resulting from the estimation error in the reference model and the measurement noise level.

These indicators are computationally cheap, and thus can be embedded. This is of particular interest in some applications, such as flutter monitoring.

Choosing the eigenvectors of matrix F as a basis for the state space of model (4) yields the following representation of the observability matrix:

$$\mathcal{O}_{p+1}(\theta) = \begin{pmatrix} \Phi \\ \Phi\Delta \\ \vdots \\ \Phi\Delta^p \end{pmatrix} \tag{13}$$

where $\Delta \triangleq \mathrm{diag}(\Lambda)$, and Λ and Φ are as in (6). Whether a nominal parameter θ_0 fits a given output covariance sequence $(R_j)_j$ is characterized by:

$$\mathcal{O}_{p+1}(\theta_0) \text{ and } \mathcal{H}_{p+1,q} \text{ have the same left kernel space.} \tag{14}$$

This property can be checked as follows. From the nominal θ_0, compute $\mathcal{O}_{p+1}(\theta_0)$ using (10), and perform e.g. a singular value decomposition (SVD) of $\mathcal{O}_{p+1}(\theta_0)$ for extracting a matrix U such that:

$$U^T\,U = I_s \text{ and } U^T\,\mathcal{O}_{p+1}(\theta_0) = 0 \tag{15}$$

Matrix U is not unique (two such matrices relate through a post-multiplication with an orthonormal matrix), but can be regarded as a function of θ_0. Then the characterization writes:

$$U(\theta_0)^T\,\mathcal{H}_{p+1,q} = 0 \tag{16}$$

3.1.2.1. Residual associated with subspace identification.

Assume now that *a reference θ_0 and a new sample $Y_1,\cdots,Y_N$ are available.* For checking whether the data agree with θ_0, the idea is to compute the empirical Hankel matrix $\widehat{\mathcal{H}}_{p+1,q}$:

$$\widehat{\mathcal{H}}_{p+1,q} \triangleq \mathrm{Hank}\left(\widehat{R}_i\right), \quad \widehat{R}_i \triangleq 1/(N-i)\sum_{k=i+1}^{N} Y_k\,Y_{k-i}^T \tag{17}$$

and to define the residual vector:

$$\zeta_N(\theta_0) \triangleq \sqrt{N}\ \mathrm{vec}\left(U(\theta_0)^T\,\widehat{\mathcal{H}}_{p+1,q}\right) \tag{18}$$

Let θ be the actual parameter value for the system which generated the new data sample, and $\mathbf{E}_\theta$ be the expectation when the actual system parameter is θ. From (13), we know that $\zeta_N(\theta_0)$ has zero mean when no change occurs in θ, and nonzero mean if a change occurs. Thus $\zeta_N(\theta_0)$ plays the role of a residual.

As in most fault detection approaches, the key issue is to design a *residual*, which is ideally close to zero under normal operation, and has low sensitivity to noises and other nuisance perturbations, but high sensitivity to small deviations, before they develop into events to be avoided (damages, faults, ...). The originality of our approach is to :

- *Design* the residual basically as a *parameter estimating function*,
- *Evaluate* the residual thanks to a kind of central limit theorem, stating that the residual is asymptotically Gaussian and reflects the presence of a deviation in the parameter vector through a change in its own mean vector, which switches from zero in the reference situation to a non-zero value.

The central limit theorem shows [48] that the residual is asymptotically Gaussian :

$$\zeta_N \xrightarrow[N\to\infty]{} \begin{cases} \mathcal{N}(0,\Sigma) & \text{under } \mathbf{P}_{\theta_0} \ , \\ \\ \mathcal{N}(\mathcal{J}\,\eta,\Sigma) & \text{under } \mathbf{P}_{\theta_0+\eta/\sqrt{N}} \ , \end{cases} \tag{19}$$

where the asymptotic covariance matrix Σ can be estimated, and manifests the deviation in the parameter vector by a change in its own mean value. Then, deciding between $\eta = 0$ and $\eta \neq 0$ amounts to compute the following χ^2-test, provided that $\mathcal{J}$ is full rank and Σ is invertible :

$$\chi^2 = \overline{\zeta}^{\,T}\,\mathbf{F}^{-1}\,\overline{\zeta} \gtrless \lambda \ . \tag{20}$$

where

$$\overline{\zeta} \triangleq \mathcal{J}^T\,\Sigma^{-1}\,\zeta_N \quad \text{and} \quad \mathbf{F} \triangleq \mathcal{J}^T\,\Sigma^{-1}\,\mathcal{J} \tag{21}$$

3.1.3. Diagnostics

A further monitoring step, often called *fault isolation*, consists in determining which (subsets of) components of the parameter vector θ have been affected by the change. Solutions for that are now described. How this relates to diagnostics is addressed afterwards.

The question: *which (subsets of) components of θ have changed ?*, can be addressed using either nuisance parameters elimination methods or a multiple hypotheses testing approach [47].

In most SHM applications, a complex physical system, characterized by a generally non identifiable parameter vector Φ has to be monitored using a simple (black-box) model characterized by an identifiable parameter vector θ. A typical example is the vibration monitoring problem for which complex finite elements models are often available but not identifiable, whereas the small number of existing sensors calls for identifying only simplified input-output (black-box) representations. In such a situation, two different diagnosis problems may arise, namely diagnosis in terms of the black-box parameter θ and diagnosis in terms of the parameter vector Φ of the underlying physical model.

The isolation methods sketched above are possible solutions to the former. Our approach to the latter diagnosis problem is basically a detection approach again, and not a (generally ill-posed) inverse problem estimation approach.

The basic idea is to note that the physical sensitivity matrix writes $\mathcal{J}\,\mathcal{J}_{\Phi\theta}$, where $\mathcal{J}_{\Phi\theta}$ is the Jacobian matrix at Φ_0 of the application $\Phi \mapsto \theta(\Phi)$, and to use the sensitivity test for the components of the parameter vector Φ. Typically this results in the following type of directional test :

$$\chi^2_\Phi = \zeta^T\,\Sigma^{-1}\,\mathcal{J}\,\mathcal{J}_{\Phi\theta}\,(\mathcal{J}^T_{\Phi\theta}\,\mathcal{J}^T\,\Sigma^{-1}\,\mathcal{J}\,\mathcal{J}_{\Phi\theta})^{-1}\,\mathcal{J}^T_{\Phi\theta}\,\mathcal{J}^T\,\Sigma^{-1}\,\zeta \gtrless \lambda \ . \tag{22}$$

It should be clear that the selection of a particular parameterization Φ for the physical model may have a non negligible influence on such type of tests, according to the numerical conditioning of the Jacobian matrices $\mathcal{J}_{\Phi\theta}$.

3.2. Thermal methods

3.2.1. Infrared thermography and heat transfer

This section introduce the infrared radiation and its link with the temperature, in the next part different measurement methods based on that principle are presented.

3.2.1.1. Infrared radiation

Infrared is an electromagnetic radiation having a wavelength between $0.2\mu m$ and 1 mm, this range begin in uv spectrum and it ends on the microwaves domain, see Figure 1.

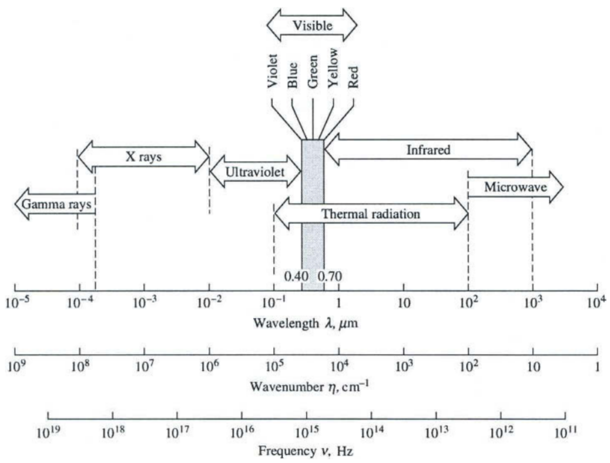


Figure 1. Electromagnetic spectrum - Credit MODEST, M.F. (1993). Radiative Heat Transfer. Academic Press.

For scientific purpose infrared can be divided in three ranges of wavelength in which the application varies, see Table 1.

Table 1. Wavelength bands in the infrared according to ISO 20473:2007

Band name	**wavelength**	**Uses \ definition**
Near infrared (PIR, IR-A, NIR)	$0.7 - 3\mu$m	Reflected solar heat flux
Mid infrared (MIR, IR-B)	$3 - 50\mu$m	Thermal infrared
Far infrared (LIR, IR-C, FIR)	$50 - 1000\mu$m	Astronomy

Our work is concentrated in the mid infrared spectral band. Keep in mind that Table 1 represents the ISO 20473 division scheme, in the literature boundaries between bands can move slightly.
The Plank's law, proposed by Max Planck en 1901, allow to compute the black body emission spectrum for various temperatures (and only temperatures), see Figure 2 left. The black body is a theoretical construction, it represents perfect energy emitter at a given temperature, cf Equation (20).

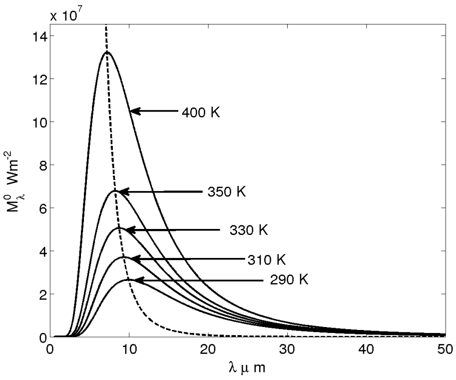
$$M^o_{\lambda,T} = \frac{C_1 \lambda^{-5}}{\exp^{\frac{C_2}{\lambda T}} - 1} \tag{23}$$

With λ the wavelength in m and T as the temperature in Kelvin. The C_1 an C_2 constant, respectively in W.m^2 and m.K are defined as follow:

$$\begin{aligned} C_1 &= 2hc^2\pi \\ C_2 &= h\frac{c}{k} \end{aligned} \tag{24}$$

with

- c The electromagnetic wave speed (in vacuum c is the light speed in m.s^{-1}).
- $k = 1.381e^{-23}$ J.K^{-1} The Boltzmann (Entropy definition from Ludwig Boltzmann 1873). It can be seen as a proportionality factor between the temperature and the energy of a system.
- $h \approx 6,62606957e^{-34}$ J.s The Plank constant. It is the link between the photons energy and their frequency.



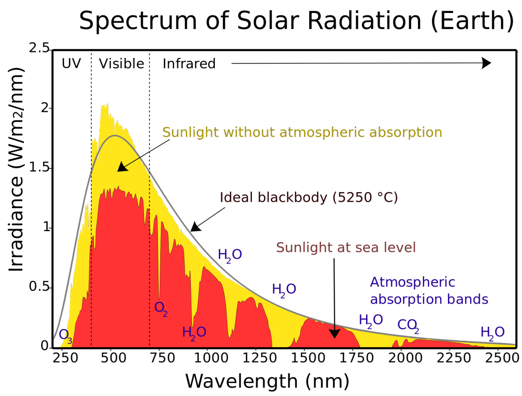


Figure 2. Left: Plank's law at various temperatures - Right: Energy spectrum of the atmosphere

By generalizing the Plank's law with the Stefan Boltzmann law (proposed first in 1879 and then in 1884 by Joseph Stefan and Ludwig Boltzmann) it is possible to address mathematically the energy spectrum of real body at each wavelength dependent of the temperature, the optical condition and the real body properties, which is the base of the infrared thermography.

For example, Figure 2 right presents the energy spectrum of the atmosphere at various levels, it can be seen that the various properties of the atmosphere affect the spectrum at various wavelengths. Other important point is that the infrared solar heat flux can be approximated by a black body at 5523,15 K.

3.2.1.2. Infrared Thermography

The infrared thermography is a way to measure the thermal radiation received from a medium. With that information about the electromagnetic flux it is possible to estimate the surface temperature of the body, see section 3.2.1.1. Various types of detector can assure the measure of the electromagnetic radiation.

Those different detectors can take various forms and/or manufacturing process. For our research purpose we use uncooled infrared camera using a matrix of microbolometers detectors. A microbolometer, as a lot of transducers, converts a radiation in electric current used to represent the physical quantity (here the heat flux).

This field of activity includes the use and the improvement of vision system, like in [3].

3.2.2. Heat transfer theory

Once the acquisition process is done, it is useful to model the heat conduction inside the cartesian domain Ω. Note that in opaque solid medium the heat conduction is the only mode of heat transfer. Proposed by Jean Baptiste Biot in 1804 and experimentally demonstrated by Joseph Fourier in 1821, the Fourier Law describes the heat flux inside a solid, cf Equation (22).

$$\varphi = k\nabla T \quad X \in \Omega \tag{25}$$

Where k is the thermal conductivity in W.m^{-1}.K ^{o}, ∇ is the gradient operator and φ is the heat flux density in Wm^{-2}. This law illustrates the first principle of thermodynamic (law of conservation of energy) and implies the second principle (irreversibility of the phenomenon), from this law it can be seen that the heat flux always goes from hot area to cold area.
An energy balance with respect to the first principle drives to the expression of the heat conduction in all point of the domain Ω, cf Equation (23). This equation has been proposed by Joseph Fourier in 1811.

$$\rho C \frac{\partial T(X,t)}{\partial t} = \nabla \cdot (k\nabla T) + P \quad X \in \Omega \tag{26}$$

With $\nabla.()$ the divergence operator, C the specific heat capacity in J.kg^{-1}.^{o}K^{-1}, ρ the volumetric mass density in kg. m^{-3}, X the space variable $X = \{x, y, z\}$ and P a possible internal heat production in W.m^{-3}.
To solve the system (23), it is necessary to express the boundaries conditions of the system. With the developments presented in section 3.2.1.1 and the Fourier's law it is possible, for example, to express the thermal radiation and the convection phenomenon which can occur at $\partial\Omega$ the system boundaries, cf Equation (24).

$$\varphi = k\nabla T \cdot n = \underbrace{h\left(T_{fluid} - T_{Boundarie}\right)}_{\text{Convection}} + \underbrace{\epsilon\sigma_s\left(T^4_{environement} - T^4_{Boundary}\right)}_{\text{Radiation}} + \varphi_0 \quad X \in \partial\Omega \tag{27}$$

Equation (24) is the so called Robin condition on the boundary $\partial\Omega$, where n is the normal, h the convective heat transfer coefficient in W.m^{-2}.K^{-1} and φ_0 an external energy contribution W.m^{-2}, in cases where the external energy contribution is artificial and controlled we call it active thermography (spotlight etc...) in the contrary it is called passive thermography (direct solar heat flux).
The systems presented in the different sections above (3.2.1 to 3.2.2) are useful to build physical models in order to represents the measured quantity. To estimate key parameters, as the conductivity, one way to do is the model inversion, the next section will introduce that principle.

3.2.3. Inverse model for parameters estimation

Lets take any model A which can for example represent the conductive heat transfer in a medium, the model is solved for a parameter vector P and it results another vector b, cf Equation (25). For example if A represents the heat transfer, b can be the temperature evolution.

$$AP = b \tag{28}$$

With A a matrix of size $n \times m$, P a vector of size m and b of size n, preferentially $n >> P$. This model is called direct model, the inverse model consist to find a vector P which satisfy the results b of the direct model. For that we need to inverse the matrix A, cf Equation (26).

$$P = A^{-1}b \tag{29}$$

Here we want find the solution AP which is closest to the acquired measures M, Equation (27).

$$AP \approx \mathcal{M} \tag{30}$$

To do that it is important to respect the well posed condition established by Jacques Hadamard in 1902

- A solution exists.
- The solution is unique.
- The solution's behavior changes continuously with the initial conditions.

Unfortunately those condition are rarely respected in our field of study. That is why we dont solve directly the system (27) but we minimise the quadratic coast function (28) which represents the Legendre-Gauss least square algorithm for linear problems.

$$min_P \left(\|AP - \mathcal{M}\|^2 \right) = min_P \left(\mathcal{F} \right) \tag{31}$$

Where $\mathcal{F}$ can be a product of matrix.

$$\mathcal{F} = [AP - \mathcal{M}]^T [AP - \mathcal{M}] \tag{32}$$

In some case the problem is still ill-posed and need to be regularized for example using the Tikhonov regularization. An elegant way to minimize the cost function $\mathcal{F}$ is compute the gradient, Equation (30) and find where it is equal to zero.

$$\nabla\mathcal{F}(P) = 2 \left[-\frac{\partial AP^T}{\partial P} \right] [AP - \mathcal{M}] = 2J(P)^T [AP - \mathcal{M}] \tag{33}$$

Where J is the sensitivity matrix of the model A to its parameter vector P.
Until now the inverse method proposed is valid only when the model A is linearly dependent of its parameter P, for the heat equation it is the case when you want to estimate the external heat flux, φ_0 in equation 24. For all the other parameters, like the conductivity k the model is non-linearly dependant of its parameter P. For such case the use of iterative algorithm is needed, for example the Levenberg-Marquardt algorithm, cf Equation (31).

$$P^{k+1} = P^k + [(J^k)^T J^k + \mu^k \Omega^k]^{-1} (J^k)^T \left[\mathcal{M} - A(P^k)\right] \tag{34}$$

Equation (31) is solved iteratively at each loop k. Some of our results with such linear or non linear method can be seen in [4] or [2], more specifically [1] is a custom implementation of the Levenberg-Marquardt algorithm based on the adjoint method (developed by Jacques Louis Lions in 1968) coupled to the conjugate gradient algorithm to estimate wide properties field in a medium.

3.3. Reflectometry-based methods for electrical engineering and for civil engineering

The fast development of electronic devices in modern engineering systems involves more and more connections through cables, and consequently, with an increasing number connexion failures. Wires and connectors are subject to ageing and degradation, sometimes under severe environmental conditions. In many applications, the reliability of electrical connexions is related to the quality of production or service, whereas in critical applications reliability becomes also a safety issue. It is thus important to design smart diagnosis systems able to detect connection defects in real time. This fact has motivated research projects on methods for fault diagnosis in this field. Some of these projects are based on techniques of reflectometry, which consist in injecting waves into a cable or a network and in analyzing the reflections, as in the example of cable hard fault diagnosis. Depending on the injected waveforms and on the methods of analysis, various techniques of reflectometry are available. They all have the common advantage of being non destructive.

At Inria the research activities on reflectometry started within the SISYPHE EPI several years ago and now continue in the I4S EPI. Our most notable contribution in this area is a method based on the *inverse scattering* theory for the computation of *distributed characteristic impedance* along a cable from reflectometry measurements [14], [11], [53]. It provides an efficient solution for the diagnosis of *soft* faults in electrical cables, like in the example illustrated in Figure 3. While most reflectometry methods for fault diagnosis are based on the detection and localization of impedance discontinuity, our method yielding the spatial profile of the characteristic impedance is particularly suitable for the diagnosis of soft faults *with no or weak impedance discontinuities*.

Fault diagnosis for wired networks have also been studied in Inria [55], [51]. The main results concern, on the one hand, simple star-shaped networks from measurements made at a single node, on the other hand, complex networks of arbitrary topological structure with complete node observations.

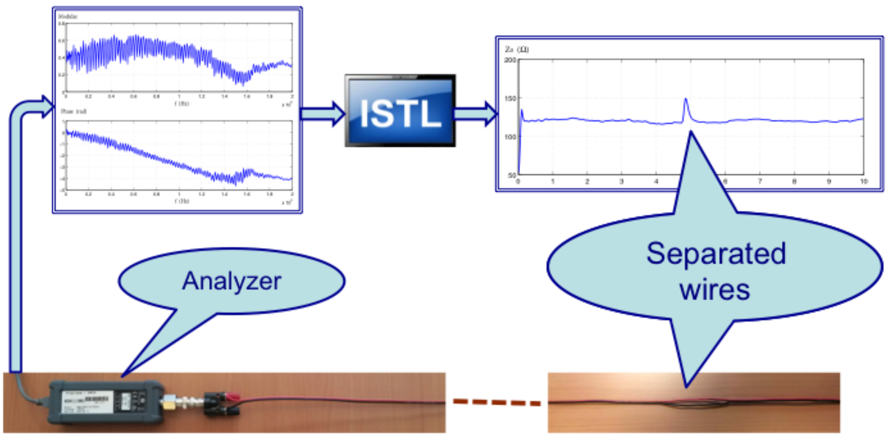


Figure 3. Inverse scattering software (ISTL) for cable soft fault diagnosis.

Though initially our studies on reflectometry were aiming at applications in electrical engineering, through our collaboration with IFSTTAR, we are also investigating applications in the field of civil engineering, by using electrical cables as sensors for monitoring changes in mechanical structures.

What follows is about some basic elements on mathematical equations of electric cables and networks, the main approach we follow in our study, and our future research directions.

3.3.1. Mathematical model of electric cables and networks

A cable excited by a signal generator can be characterized by the telegrapher's equations [52]

$$\begin{aligned} \frac{\partial}{\partial z} V(t,z) + L(z)\frac{\partial}{\partial t} I(t,z) + R(z) I(t,z) = 0 \\ \frac{\partial}{\partial z} I(t,z) + C(z)\frac{\partial}{\partial t} V(t,z) + G(z) V(t,z) = 0 \end{aligned} \tag{35}$$

where t represents the time, z is the longitudinal coordinate along the cable, $V(t,z)$ and $I(t,z)$ are respectively the voltage and the current in the cable at the time instant t and at the position z, $R(z), L(z), C(z)$ and $G(z)$ denote respectively the series resistance, the inductance, the capacitance and the shunt conductance per unit length of the cable at the position z. The left end of the cable (corresponding to $z = a$) is connected to a voltage source $V_s(t)$ with internal impedance R_s. The quantities $V_s(t)$, R_s, $V(t,a)$ and $I(t,a)$ are related by

$$V(t,a) = V_s(t) - R_s I(t,a). \tag{36}$$

At the right end of the cable (corresponding to $z = b$), the cable is connected to a load of impedance R_L, such that

$$V(t,b) \quad = R_L I(t,b). \tag{37}$$

One way for deriving the above model is to spatially discretize the cable and to characterize each small segment with 4 basic lumped parameter elements for the j-th segment: a resistance ΔR_j, an inductance ΔL_j, a capacitance ΔC_j and a conductance ΔG_j. The entire circuit is described by a system of ordinary differential equations. When the spatial discretization step size tends to zero, the limiting model leads to the telegrapher's equations (32).

A wired network is a set of cables connected at some nodes, where loads and sources can also be connected. Within each cable the current and voltage satisfy the telegrapher's equations (32), whereas at each node the current and voltage satisfy the Kirchhoff's laws, unless in case of connector failures.

3.3.2. The inverse scattering theory applied to cables

The inverse scattering transform was developed during the 1970s-1980s for the analysis of some nonlinear partial differential equations [50]. The visionary idea of applying this theory to solving the cable inverse problem goes also back to the 1980s [49]. After having completed some theoretic results directly linked to practice [14], [53], we started to successfully apply the inverse scattering theory to cable soft fault diagnosis, in collaboration with GEEPS-SUPELEC [11].

To link electric cables to the inverse scattering theory, the telegrapher's equations (32) are transformed in a few steps to fit into a particular form studied in the inverse scattering theory. The Fourier transform is first applied to transform the time domain model (32) into the frequency domain, the spatial coordinate z is then replaced by the propagation time

$$x(z) = \int_0^z \sqrt{L(s)C(s)}ds$$

and the frequency domain variables $V(\omega,x), I(\omega,x)$ are replaced by the pair

$$\begin{aligned} \nu_1(\omega,x) = \frac{1}{2}\left[Z_0^{-\frac{1}{2}}(x)U(\omega,x) - Z_0^{\frac{1}{2}}(x)I(\omega,x)\right] \\ \nu_2(\omega,x) = \frac{1}{2}\left[Z_0^{-\frac{1}{2}}(x)U(\omega,x) + Z_0^{\frac{1}{2}}(x)I(\omega,x)\right] \end{aligned} \tag{38}$$

with

$$Z_0(x) = \sqrt{\frac{L(x)}{C(x)}}. \tag{39}$$

These transformations lead to the Zakharov-Shabat equations

$$\begin{aligned}\frac{d\nu_1(\omega,x)}{dx} + ik\nu_1(\omega,x) = q^*(x)\nu_1(\omega,x) + q^+(x)\nu_2(\omega,x)\\ \frac{d\nu_2(\omega,x)}{dx} - ik\nu_2(\omega,x) = q^-(x)\nu_1(\omega,x) - q^*(x)\nu_2(\omega,x)\end{aligned} \tag{40}$$

with

$$\begin{aligned} q^{\pm}(x) &= -\frac{1}{4}\frac{d}{dx}\left[ln\frac{L(x)}{C(x)}\right] \mp \frac{1}{2}\left[\frac{R(x)}{L(x)} - \frac{G(x)}{C(x)}\right]\\ &= -\frac{1}{2Z_0(x)}\frac{d}{dx}Z_0(x) \mp \frac{1}{2}\left[\frac{R(x)}{L(x)} - \frac{G(x)}{C(x)}\right]\\ q^*(x) &= \frac{1}{2}\left[\frac{R(x)}{L(x)} + \frac{G(x)}{C(x)}\right].\end{aligned} \tag{41}$$

These equations have been well studied in the inverse scattering theory, for the purpose of determining partly the “potential functions” $q^{\pm}(x)$ and $q^*(x)$ from the scattering data matrix, which turns out to correspond to the data typically collected with reflectometry instruments. For instance, it is possible to compute the function $Z_0(x)$ defined in (36), often known as the characteristic impedance, from the reflection coefficient measured at one end of the cable. Such an example is illustrated in Figure 3. Any fault affecting the characteristic impedance, like in the example of Figure 3 caused by a slight geometric deformation, can thus be efficiently detected, localized and characterized.

3.4. Research Program

The research will first focus on the extension and implementation of current techniques as developed in I4S and IFSTTAR. Before doing any temperature rejection on large scale structures as planned, we need to develop good and accurate models of thermal fields. We also need to develop robust and efficient versions of our algorithms, mainly the subspace algorithms before envisioning linking them with physical models. Briefly, we need to mature our statistical toolset as well as our physical modeling before mixing them together later on.

3.4.1. Vibration analysis and monitoring

3.4.1.1. Direct vibration modeling under temperature changes

This task builds upon what has been achieved in the CONSTRUCTIF project, where a simple formulation of the temperature effect has been exhibited, based on relatively simple assumptions. The next step is to generalize this modeling to a realistic large structure under complex thermal changes. Practically, temperature and resulting structural prestress and pre strains of thermal origin are not uniform and civil structures are complex. This leads to a fully 3D temperature field, not just a single value. Inertia effects also forbid a trivial prediction of the temperature based on current sensor outputs while ignoring past data. On the other side, the temperature is seen as a nuisance. That implies that any damage detection procedure has first to correct the temperature effect prior to any detection.

Modeling vibrations of structures under thermal prestress does and will play an important role in the static correction of kinematic measurements, in health monitoring methods based on vibration analysis as well as in durability and in the active or semi-active control of civil structures that by nature are operated under changing environmental conditions. As a matter of fact, using temperature and dynamic models the project aims at correcting the current vibration state from induced temperature effects, such that damage detection algorithms rely on a comparison of this thermally corrected current vibration state with a reference state computed or measured at a reference temperature. This approach is expected to cure damage detection algorithms from the environmental variations.

I4S will explore various ways of implementing this concept, notably within the FUI SIPRIS project.

3.4.1.2. Damage localization algorithms (in the case of localized damages such as cracks)

During the CONSTRUCTIF project, both feasibility and efficiency of some damage detection and localization algorithms were proved. Those methods are based on the tight coupling of statistical algorithms with finite element models. It has been shown that effective localization of some damaged elements was possible, and this was validated on a numerical simulated bridge deck model. Still, this approach has to be validated on real structures.

On the other side, new localization algorithms are currently investigated such as the one developed conjointly with University of Boston and tested within the framework of FP7 ISMS project. These algorithms will be implemented and tested on the PEGASE platform as well as all our toolset.

When possible, link with temperature rejection will be done along the lines of what has been achieved in the CONSTRUCTIF project.

3.4.1.3. Uncertainty quantification for system identification algorithms

Some emphasis will be put on expressing confidence intervals for system identification. It is a primary goal to take into account the uncertainty within the identification procedure, using either identification algorithms derivations or damage detection principles. Such algorithms are critical for both civil and aeronautical structures monitoring. It has been shown that confidence intervals for estimation parameters can theoretically be related to the damage detection techniques and should be computed as a function of the Fisher information matrix associated to the damage detection test. Based on those assumptions, it should be possible to obtain confidence intervals for a large class of estimates, from damping to finite elements models. Uncertainty considerations are also deeply investigated in collaboration with Dassault Aviation in Mellinger PhD thesis or with Northeastern University, Boston, within Gallegos PhD thesis.

3.4.2. Reflectometry-based methods for civil engineering structure health monitoring

The inverse scattering method we developed is efficient for the diagnosis of all soft faults affecting the characteristic impedance, the major parameter of a cable. In some particular applications, however, faults would rather affect the series resistance (ohmic loss) or shunt conductance (leakage loss) than the characteristic impedance. The first method we developed for the diagnosis of such losses had some numerical stability problems. The new method [46], [26] is much more reliable and efficient. It is also important to develop efficient solutions for long cables, up to a few kilometers.

For wired networks, the methods we already developed cover either the case of simple networks with a single node measurement or the case of complex networks with complete node measurements. Further developments are still necessary for intermediate situations.

In terms of applications, the use of electric cables as sensors for the monitoring of various structures is still at its beginning. We believe that this new technology has a strong potential in different fields, notably in civil engineering and in materials engineering.

3.4.3. Non Destructive testing of CFRP bonded on concrete through active thermography

Strengthening or retrofitting of reinforced concrete structures by externally bonded fibre-reinforced polymer (FRP) systems is now a commonly accepted and widespread technique. However, the use of bonding techniques always implies following rigorous installation procedures. The number of carbon fibre-reinforced polymer (CFRP) sheets and the glue layer thickness are designed by civil engineers to address strengthening objectives. Moreover, professional crews have to be trained accordingly in order to ensure the durability and long-term performance of the FRP reinforcements. Conformity checking through an 'in situ' verification of the bonded FRP systems is then highly desirable. The quality control programme should involve a set of adequate inspections and tests. Visual inspection and acoustic sounding (hammer tap) are commonly used to detect delaminations (disbonds). Nevertheless, these techniques are unable to provide sufficient information about the depth (in case of multilayered composite) and width of the disbonded areas. They are also incapable of evaluating the degree of adhesion between the FRP and the substrate (partial delamination, damage of the resin and poor mechanical properties of the resin). Consequently, rapid and efficient inspection methods are required. Among the non-destructive (NDT) methods currently under study, active infrared thermography is investigated due to its ability to be used in the field. In such context and to reach the aim of having an in situ efficient NDT method, we carried out experiments and subsequent data analysis using thermal excitation. Image processing, inverse thermal modelling and 3D numerical simulations are used and then applied to experimental data obtained in laboratory conditions.

3.4.4. IRSHM: Multi-Sensing system for outdoor thermal monitoring

Ageing of transport infrastructures combined with traffic and climatic solicitations contribute to the reduction of their performances. To address and quantify the resilience of civil engineering structure, investigations on robust, fast and efficient methods are required. Among research works carried out at IFSTTAR, methods for long term monitoring face an increasing demand. Such works take benefits of this last decade technological progresses in ICT domain.

Thanks to IFSTTAR years of experience in large scale civil engineering experiment, I4S is able to perform very long term thermal monitoring of structures exposed to environmental condition, as the solar heat flux, natural convection or seasonal perturbation. Informations system are developed to asses the data acquisition and researchers work on the quantification of the data to detect flaws emergence on structure, those techniques are also used to diagnose thermal insulation of buildings or monitoring of guided transport infrastructures, Figure 4 left. Experiments are carried out on a real transport infrastructure open to traffic and buildings. The detection of the inner structure of the deck is achieved by image processing techniques (as FFT), principal component thermography (PCT), Figure 4 right, or characterization of the inner structure thanks to an original image processing approach.

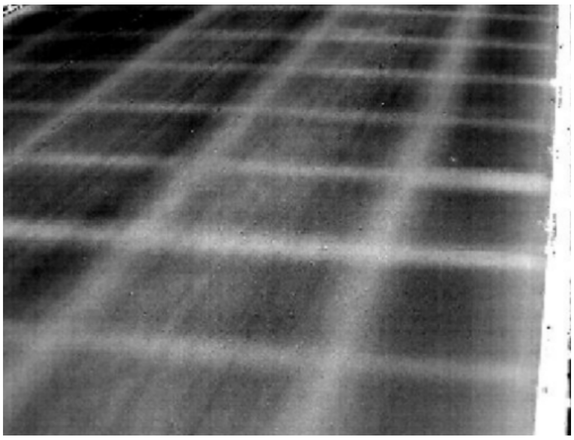


Figure 4. Left: Image in the visible spectrum of the deck surface - Right: PCT result on a bridge deck

For the next few years, I4S is actively implied in the SenseCity EQUIPEX (http://sense-city.ifsttar.fr/) where our informations systems are used to monitor a mini-city replica, Figure 5.

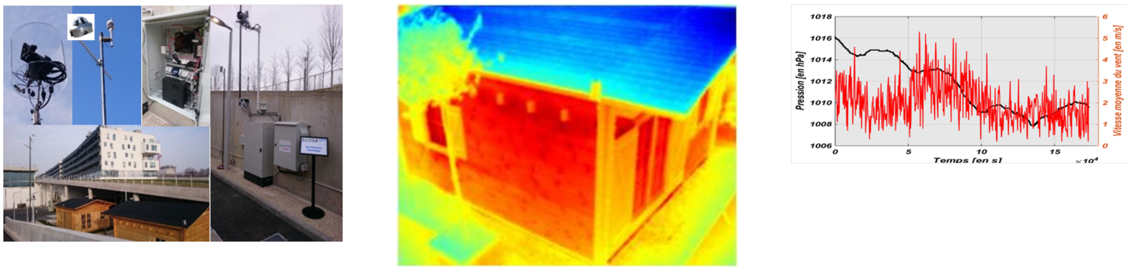


Figure 5. Various view and results of the SenseCity experimentation site - (site and hardware view, IR imaging, Environmental Monitoring)

3.4.5. R5G: The 5th Generation Road

The road has to reinvent itself periodically in response to innovations, societal issues and rising user expectations. The 5th Generation Road (R5G) focuses firmly on the future and sets out to be automated, safe, sustainable and suited to travel needs. Several research teams are involved in work related to this flagship project for IFSTTAR, which is a stakeholder in the Forever Open Road. Through its partnership with the COSYS (IFSTTAR) department, I4S is fully implicated in the development of the 5th Generation Road.

Most of the innovations featured in R5G are now mature, for example communication and few solutions for energy exchange between the infrastructure, the vehicle and the network manager; recyclable materials with the potential for self-diagnosis and repair, a pavement surface that remains permanently optimal irrespective of climatic variations... Nevertheless, implementing them on an industrial scale at a reasonable cost still represents a real challenge. Consultation with the stakeholders (researchers, industry, road network owners and users) has already established the priorities for the creation of full-scale demonstrators. The next stages are to achieve synergy between the technologies tested by the demonstrators, to manage the interfaces and get society to adopt R5G.

4. Application Domains

4.1. Civil Engineering

For at least three decades, monitoring the integrity of the civil infrastructure has been an active research topic because of major economical and societal issues, such as durability and safety of infrastructures, buildings and networks. Control of civil structures began a century ago. At stake is the mastering of the ageing of the bridges, as in America (US, Canada) and Great Britain, or the resistance to seismic events and the protection of the cultural heritage, as in Italy and Greece. The research effort in France is very ancient since for example early developments of optical methods to monitor civil structures began in the 70s and SHM practice can be traced back to the 50s with the vibrating wire sensors as strain gauges for dams. Stille the number of sensors actually placed on civil structures is kept to a minimum, mainly for cost reasons, but also because the return on investment sensing and data processing technologies is not properly established for civil structures. One of the current thematic priorities of the C2D2 governmental initiative is devoted to construction monitoring and diagnostics. The picture in Asia (Japan, and also China) is somewhat different, in that recent or currently built bridges are equipped with hundreds if not thousands of sensors, in particular the Hong Kong-Shenzen Western Corridor and Stonecutter Bridge projects. However, the actual use of available data for operational purpose remains unclear.

Among the challenges for vibration-based bridges health monitoring, two major issues are the different kinds of (non measured) excitation sources and the environmental effects. Typically the traffic on *and* under the bridge, the wind and also the rain, contribute to excite the structure, and influence the measured dynamics. Moreover, the temperature is also known to affect the eigenfrequencies and mode-shapes, to an extent which can be significant w.r.t. the deviations to be monitored.

Thermomechanical prestress states affect the dynamic and the static behavior of most bridges, not only of very long and flexible ones. So, the reliable and fast determination of the state of prestress and prestrain associated with a temperature field becomes a crucial step in several engineering processes such as the health monitoring of civil structures. The best possible reconstruction of the temperature field could then become part of a complete process including massively distributed sensing of thermomechanical information on the structure, modeling and algorithms for the on-line detection of damages in the sense of abnormalities with regard to a nominal state, the whole chain being encapsulated in professional tools used by engineers in charge of real-life structural monitoring. For lack of an adequate mobilization of the useful multidisciplinary skills, this way remains about unexplored today.

4.2. Electrical cable and network monitoring

The fast development of electronic devices in modern engineering systems comes with more and more connections through cables, and consequently, the reliability of electric connections becomes a crucial issue. For example, in a modern automotive vehicle, the total length of onboard cables has tremendously increased during the last decades and is now up to 4km. These wires and connectors are subject to ageing or degradation because of severe environmental conditions. In this area, reliability becomes a safety issue. In some other domains, cable defects may have catastrophic consequences. It is thus a crucial challenge to design smart embedded diagnosis systems able to detect wired connection defects in real time. This fact has motivated research projects on methods for fault diagnosis in electric transmission lines and wired networks. Original methods have been recently developed by Inria, notably based on the inverse scattering theory, for cable and network monitoring. Further developments concern both theoretic study and industrial applications.

4.3. Aeronautics

Improved safety and performance and reduced aircraft development and operating costs are major concerns in aeronautics industry. One critical design objective is to clear the aircraft from unstable aero-elastic vibrations (flutter) in all flight conditions. Opening of flight domain requires a careful exploration of the dynamical behavior of the structure subject to vibration and aero-servo-elastic forces. This is achieved via a combination of ground vibration tests and in flight tests. For both types of tests, various sensors data are recorded, and modal analyses are performed. Important challenges of the in-flight modal analyses are the limited choices for measured excitation inputs, and the presence of unmeasured natural excitation inputs (turbulence). Today, structural flight tests require controlled excitation by ailerons or other devices, stationary flight conditions (constant elevation and speed), and no turbulence. As a consequence, flight domain opening requires a lot of test flights and its costly. This is even worse for aircrafts having a large number of variants (business jets, military aircrafts). A key challenge is therefore to allow for exploiting more data under more conditions during flight tests: uncontrolled excitation, nonstationary conditions.

5. Highlights of the Year

5.1. Highlights of the Year

- In 2016, uncertainty quantification for modal analysis has been transferred to ARTeMIS software http://www.svibs.com/newsletter/newsletter_2016_09.aspx.
- In 2016, a patent has been filed by N. Berrabah and Q. Zhang, jointly with EDF and Inria [46].

- PEDAL-LORA monitoring sensor has been awarded by the European Railway Cluster Price in railway innovation.

6. New Software and Platforms

6.1. Cloud2IR

KEYWORDS: Sensors - SHM (Structural Health Monitoring) - Sensors network - Open Geospatial Consortium - OGC

SCIENTIFIC DESCRIPTION

Cloud2IR is an helpful tool to build physicals models for the SHM of civil engineering structures.

In this context the software is deployed in the SenseCity-EQUIPEX driven by IFSTTAR (http://sense-city.ifsttar.fr/)

FUNCTIONAL DESCRIPTION

Cloud 2IR is a software dedicated to the structural health monitoring of civil engineering structures thanks to long term thermal imaging. Its particularity lies in the fact that it is based on a generic approach of the acquisition system concept and the format of the data. That allow it to apply to other types of sensor.

- Partner: IFSTTAR
- Contact: Jean Dumoulin

This work has been developed during the ADT of Antoine Crinière, Cloud2SM.

6.2. PEGASE

KEYWORD: Sensors - SHM (Structural Health Monitoring)

SCIENTIFIC DESCRIPTION

PEGASE (Plate-forme Experte Générique pour Applications Sans-fil Embarquées) is a generic and high level wireless sensor platform. Currently, the setup of the new PEGASE 2 platform is finalized as the technological successor of the previous PEGASE platform developed by IFSTTAR. This PEGASE 1 platform is licensed and disseminated by the third party company A3IP since 2008 and has been sold in thousands of units.

Based on various feedback from monitoring applications of PEGASE, and due to the fast obsolescence of electronic devices, the design of the new PEGASE 2 platform has been launched in 2013. Some of the main functions of PEGASE are reinforced:

- Software genericity: use of a Linux embedded OS to make application development independent from the hardware, and to enable the user to manage the system without any physical and heavy operations.
- Hardware genericity: with a principle of daughter and mother boards, each redundant need is embedded (processing, memory, timing, GPS, energy, etc) where each pluggable daughter board implements a specific function (sensing, 3G, Ethernet, communication, signal processing and relay control).
- Accurate time synchronization principle: based on an original GPS and PPS algorithm, PEGASE platform is one of the only boards able to time-stamp data from sensors or any event with high accuracy and in a deterministic way.

On PEGASE 2 platform, previous principles are maintained or extended. A full electronic design from scratch has been decided by the team in 2014 to maximize its capacities in terms efficiency, cost, energy consumption, etc. The main new characteristics of PEGASE 2 are:

- A “real” Linux kernel is now flashed inside the board based on a professional Debian 4.0 (or higher version) of Linux. This Linux branch is the only one validated by IEEE to suit embedded applications
- Previous SDK in C language has been ported and improved in C++ with a generalization of the signal/slot principle to offer end-users a full programing context based on “event driven developments”.
- Extended hardware capacities: GigaBytes of memory, multiple USB port, etc...
- PEGASE 2 board implements new hardware such as a on-board Battery Management System (BMS) to be able to manage an energy efficiency based on a battery and a solar cell; PEGASE 2 mother board integrates a MEMS Mpu9150 from Invenses company wich is a MEMS that provides Acceleration, Temperature and Gyroscope in 3D,...

In 2015 and 2016 various functional daughter boards have been designed to complete the PEGASE 2 panoply:

- 8 analog and 8 digital daughter boards
- LORA protocol Daughter Board
- 3G Daughter Board

Figure 6. PEGASE and PEGASE daughter board

In 2016, based on an industrial contract with the company SDEL-CC (subsidiary company of Vinci group) an important algorithm has been implemented in PEGASE 2 board to make it able to time stamp any physical event up to 10 nanoseconds (independently from the wireless protocol or the distances between the platform that could be some tens of kilometers). This time stamping ability is unique and makes PEGASE 2 the only wireless device in the world with such a time accuracy in a deterministic way and in universal time (based on its GSP/PPS principle).

The most significant development in the PEGASE 2 context is the development of a Generic Cloud Server Application:

- manage multiple instrumentation projects
- various sensors (based on PEGASE 1 or 2 or others) can be set up
- collect big data based on Mongo DB database
- visualize the graphs of data from sensors in f(t) on dynamic charts
- export data to files, through specific API, or to third-party software (Matlab, QT C++...)

In the last two years the Generic Supervisor became a professional software product that is transfered under industrial Licensing, e.g. to the companies Power Lan and Stimio. By the end of 2016, at least these 2 companies will be officially licensed by IFSTTAR to disseminate the Supervisor.

- Participants: Vincent Le Cam, Mathieu Le Pen, Laurent Mevel and Michael Doehler
- Contact: Michael Doehler
- URL: http://www.a3ip.com/joomla/index.php?option=com_content&view=article&id=12&Itemid=8

Figure 7. Supervisor

6.3. TrackingMecaSys

KEYWORDS: Bayesian estimation - Monte-Carlo - GPGPU - Kalman filter - Particular filter - Vibrating system

FUNCTIONAL DESCRIPTION

Implementation of a method based on the use of Bayesian modal parameter recursive estimation based on a particular Kalman filter algorithm with decoupled distributions for mass and stiffness. Algorithm optimized for a GPGPU implementation. This work has been done during ADT of Antoine Crinière and will be updated during the postdoc of S. Sen.

- Contact: Laurent Mevel

7. New Results

7.1. Outdoor InfraRed Thermography

7.1.1. Autonomous software architecture standardized for infrared and environmental SHM : Cloud2IR

Participants: Antoine Crinière, Jean Dumoulin, Laurent Mevel.

Cloud2IR is an autonomous software architecture, allowing multi-sensor connection (i.e. Infrared Thermography), dedicated to the long term monitoring of infrastructures. Past experimentations have shown the need as well as usefulness of such system. The system has been developed in order to cut down software integration time which facilitates the system adaptation to each experiment specificity. That is why we propose a bi-headed architecture. A specialized part, it represents the sensor specific development as well as their drivers and their different fixed configurations. In our case, as infrared camera are slightly different than other kind of sensors, the system implement in addition an RTSP server which can be used to set up the FOV as well as other measurement parameter considerations and a generic part, which can be seen as the data management side. This last part can be seen as the first embryo of a future generic framework dedicated to the data management of local multisensors (DaMaLoS). It is able to aggregate any sensor data, type or size and automatically encapsulate them in various generic data format as HDF5 or cloud data as OGC SWE standard. This whole part is also responsible of the acquisition scenario the local storage management and the network management through SFTP or SOAP for OGC Web services. Cloud2IR has been deployed on field since more than one year at the SenseCity outdoor test bed and several month at the Inria test bed, both located in France. The system aggregates various sensors as infrared camera, a GPS, multiple pyranometers, a weather station and a proprietary access to the SenseCity data viewer.[40][41]

7.1.2. GPU Improved quantitative analysis of Longterm Infrared-Thermography Data

Participants: Antoine Crinière, Jean Dumoulin, Laurent Mevel.

Since the past decade, infrared thermography coupled with inverse models based on 1d thermal quadrupoles have shown their usefulness in civil engineering by first showing their ability to assess the quantitative non destructive testing of concrete repaired by bonded CFRP plate over a wide area (i.e. repaired or reinforced concrete beams). On the other hand early implementations of long terms monitoring methods based on such approach have given their first results over a whole bridge deck. The experimental method, allow us to have the apparent surface temperature field evolution with time for a wide area divided in pixels. Knowing this specificity, the procedure aims to apply an independent model to each pixel in order to retrieve physical properties map. Such treatment can have a high computational cost. We propose various improvement of our procedure based on GPGPU paradigm in order to shorten the computational time. This study will detail an experimental procedure able to assess the long term thermal monitoring of a bridge deck over days and to draw properties maps of the inner structure. [28]

7.1.3. Infrared thermography for cultural heritage monitoring

Participant: Jean Dumoulin.

Radiation theory helps us to introduce infrared thermography. Infrared thermography is first presented in its passive mode and followed by considerations on active mode. Some processing analysis approaches are described. They belong to signal and image processing domain or to heat transfer domain. Illustration of results obtained with such analysis approaches are described on two experiments carried out in quasi laboratory conditions. Then, a case case study of the monitoring of the Viaduct Basento in Potenza (Southern Italy) is presented. Two features make fascinating this case study. The first one regards the fact that Viaduct Basento is probably the most important and visionary architectural work of the famous structural engineer Sergio Musmeci. The second aspect concerns the application, almost unique in the scientific literature, of an integrated diagnosis approaches combining a wide set of electromagnetic sensing technologies combined with advanced civil engineering analysis methodologies and tools.[44] [45] [22] [23]

7.2. Smarts roads and R5G

7.2.1. Positive surface temperature pavement

Participants: Jean Dumoulin, Nicolas Le Touz.

The mobility during winter season in France mainly relies on the use of de-icers, with an amount ranging from two hundreds thousands tons up to two millions tons for the roads only. Besides the economic impact, there are many concerns on their environmental and infrastructure, both on roads and on airports. In such context and in the framework of the R5G (5th Generation Road) project driven by IFSTTAR, investigations were carried out on the way to modify the infrastructure to maintain pavement surface at a temperature above water freezing point. Two distinct approaches, that can could be combined, were selected. The first one consisted in having a heated fluid circulating in a porous layer within an asphalt concrete pavement sample. The second one specifically relied on the use of paraffin phase change materials (PCM) in cement concrete pavement ones. Experiments on enhanced pavement samples were conducted in a climatic chamber to simulate winter conditions for several continuous days, including wind and precipitations, and monitored by infrared thermography. [24]

7.3. Methods for building performance assessment

7.3.1. Building performance assessment

Participants: Jordan Brouns, Jean Dumoulin, Alexandre Nassiopoulos, Nicolas Le Touz.

Accurate building performance assessment is necessary for the design of efficient energy retrofit operations and to foster the development of energy performance contracts. An important barrier however is that simulation tools fail to accurately predict the actual energy consumption. Two methodology are adressed, first combining thermal sensor output and inverse algorithms to determine the key parameters of a multizone thermal model [15] then assessing wall thermal resistance estimation using infrared thermography and microwave coupling [38][34][43]

7.4. System identification

7.4.1. Variance estimation of modal parameters from subspace-based system identification

Participants: Michael Doehler, Laurent Mevel.

This work has been carried out in collaboration with Philippe Mellinger (former PhD student with Dassault Aviation, now CEA).

An important step in the operational modal analysis of a structure is to infer on its dynamic behavior through its modal parameters. When output-only data is available, i.e. measured responses of the structure, frequencies, damping ratios and mode shapes can be identified assuming that ambient sources like wind or traffic excite the system sufficiently. When also input data is available, i.e. signals used to excite the structure, input/output identification algorithms are used. The use of input information usually provides better modal estimates in a desired frequency range. When identifying the modal parameters from noisy measurement data, the information on their uncertainty is most relevant. In this work, new variance computation schemes for modal parameters are developed for four subspace algorithms, including output-only and input/output methods, as well as data-driven and covariance-driven methods. For the input/output methods, the known inputs are considered as realizations of a stochastic process. Based on Monte Carlo validations, the quality of identification, accuracy of variance estimations and sensor noise robustness are discussed. Finally these algorithms are applied on real measured data obtained during vibrations tests of an aircraft. [19] [37]

7.4.2. Bayesian parameter estimation for parameter varying systems using interacting Kalman filters

Participants: Antoine Crinière, Laurent Mevel, Jean Dumoulin.

Method based on the use of Bayesian modal parameter recursive estimation based on a particular Kalman filter algorithm with decoupled distributions for mass and stiffness. Particular Kalman filtering is a combination of two widely used Bayesian estimation methods working together: the particle filter (also called sequential Monte Carlo samplings) and the Kalman filter. Usual system identification techniques for civil and mechanical structures assume the availability of large set of data derived from a stationary quasi steady structure. On the opposite, several scenarios involve time varying structures. For example, due to interaction with aerodynamics in aeronautics, some critical parameter may have to be monitored, for instability monitoring (leading possibly to flutter) of in flight data due to fuel consumption and speed change. This relates to the monitoring of time varying structural parameters such as frequencies and damping ratios. The main idea of a particular Kalman filter is to consider stochastic particles evolving in the parameter space. For each particle, a corresponding linear state is recursively estimated by applying a Kalman filter to the mechanical system, whose modal parameters are driven by the evolution of this time-varying particle. In order to provide fast and convincing results for large time varying structure, such as an airplane, the execution time of the method has to be improved. Within the Cloud2sm ADT a GPGPU implementation of the algorithm have been developed, now a post-doctoral position have been obtained to improve the algorithm reliability.[29]

7.4.3. Stability of the Kalman filter for continuous time output error systems

Participant: Qinghua Zhang.

This work has been carried out in collaboration with Boyi Ni (SAP Labs China).

The stability of the Kalman filter is usually ensured by the uniform complete controllability *regarding the process noise* and the uniform complete observability of linear time varying systems. This work studies the case of continuous time *output error* systems, in which the process noise is totally absent. The classical stability analysis assuming the controllability regarding the process noise is thus not applicable. It is shown in this work that the uniform complete observability *alone* is sufficient to ensure the asymptotic stability of the Kalman filter applied to time varying *output error* systems, regardless of the stability of the considered systems themselves. The exponential or polynomial convergence of the Kalman filter is then further analyzed for particular cases of stable or unstable output error systems. The results of this work have been published in [20].

7.4.4. Parameter uncertainties quantification for finite element based subspace fitting approaches

Participants: Guillaume Gautier, Laurent Mevel, Michael Doehler.

This work has been carried out in collaboration with Jean-Mathieu Mencik and Roger Serra (INSA Centre Val de Loire).

We address the issue of quantifying uncertainty bounds when updating the finite element model of a mechanical structure from measurement data. The problem arises as to assess the validity of the parameters identification and the accuracy of the results obtained. A covariance estimation procedure is proposed about the updated parameters of a finite element model, which propagates the data-related covariance to the parameters by considering a first-order sensitivity analysis. In particular, this propagation is performed through each iteration step of the updating minimization problem, by taking into account the covariance between the updated parameters and the data-related quantities. Numerical simulations on a beam show the feasibility and the effectiveness of the method. [31]

7.4.5. Embedded subspace-based modal analysis and uncertainty quantification

Participants: Vincent Le Cam, Michael Doehler, Mathieu Le Pen, Ivan Guéguen, Laurent Mevel.

Operational modal analysis is an important step in many methods for vibration-based structural health monitoring. These methods provide the modal parameters (frequencies, damping ratios and mode shapes) of the structure and can be used for monitoring over time. For a continuous monitoring the excitation of a structure is usually ambient, thus unknown and assumed to be noise. Hence, all estimates from the vibration measurements are realizations of random variables with inherent uncertainty due to unknown excitation, measurement noise and finite data length. Estimating the standard deviation of the modal parameters on the same dataset offers significant information on the accuracy and reliability of the modal parameter estimates. However, computational and memory usage of such algorithms are heavy even on standard PC systems in Matlab, where reasonable computational power is provided. In this work, we examine an implementation of the covariance-driven stochastic subspace identification on the wireless sensor platform PEGASE, where computational power and memory are limited. Special care is taken for computational efficiency and low memory usage for an on-board implementation, where all numerical operations are optimized. The approach is validated from an engineering point of view in all its steps, using simulations and field data from a highway road sign structure. [33]

7.5. Damage diagnosis

7.5.1. Estimation of distributed and lumped ohmic losses in electrical cables

Participants: Nassif Berrabah, Qinghua Zhang.

This work has been carried out in the framework of a CIFRE PhD project in collaboration with EDF R&D.

Cables play an important role in modern engineering systems, from power transmission to data communication. In order to ensure reliable and cost-efficient operations, as well as a high level of performance, efficient tools are needed to assess and monitor cables. Hard faults are well handled by existing techniques, whereas soft fault diagnosis still represents an important challenge for current researches. This work focuses on the detection, localization, and estimation of resistive soft fault in electrical cables from reflectometry measurements. A method for the computation of the distributed resistance profile along the cable under test has been developed. Both experimental and simulation results confirm its effectiveness, as reported in the conference paper [26]. A patent based on this work has been registered at INPI (see Section 10.1.4.1).

7.5.2. Fault detection, isolation and quantification from Gaussian residuals

Participants: Michael Doehler, Laurent Mevel, Qinghua Zhang.

Despite the general acknowledgment in the Fault Detection and Isolation (FDI) literature that FDI are typically accomplished in two steps, namely residual generation and residual evaluation, the second step is by far less studied than the first one. This work investigates the residual evaluation method based on the local approach to change detection and on statistical tests. The local approach has the remarkable ability of transforming quite general residuals with unknown or non Gaussian probability distributions into a standard Gaussian framework, thanks to a central limit theorem. In this work, the ability of the local approach for fault quantification is exhibited, whereas previously it was only presented for fault detection and isolation. The numerical computation of statistical tests in the Gaussian framework is also revisited to improve numerical efficiency. An example of vibration-based structural damage diagnosis is presented to motivate the study and to illustrate the performance of the proposed method. [17]

7.5.3. Performance of damage detection in dependence of sample length and measurement noise

Participants: Saeid Allahdadian, Michael Doehler, Laurent Mevel.

In this work the effects of measuring noise and number of samples is studied on the stochastic subspace damage detection (SSDD) technique. In previous studies, the effect of these practical parameters was examined on simulated measurements from a model of a real structure. In this study, these effects are formulated for the expected damage index evaluated from a Chi-square distributed value. Several theorems that describe the effects are proposed and proved. These theorems are used to develop a guideline to serve the user of the SSDD method to face these effects. [25]

7.5.4. Statistical damage localization with stochastic load vectors

Participants: Md Delwar Hossain Bhuyan, Michael Doehler, Laurent Mevel.

The Stochastic Dynamic Damage Locating Vector (SDDLV) method is an output-only damage localization method based on both a Finite Element (FE) model of the structure and modal parameters estimated from output-only measurements in the damage and reference states of the system. A vector is obtained in the null space of the changes in the transfer matrix computed in both states and then applied as a load vector to the model. The damage location is related to this stress where it is close to zero. In previous works an important theoretical limitation was that the number of modes used in the computation of the transfer function could not be higher than the number of sensors located on the structure. It would be nonetheless desirable not to discard information from the identification procedure. In this work, the SDDLV method has been extended with a joint statistical approach for multiple mode sets, overcoming this restriction on the number of modes. The new approach is validated in a numerical application, where the outcomes for multiple mode sets are compared with a single mode set. From these results, it can be seen that the success rate of finding the correct damage localization is increased when using multiple mode sets instead of a single mode set. [27]

7.5.5. Classification of vibration-based damage localization methods

Participant: Michael Doehler.

This work, issued from the COST Action TU1402, is in collaboration with M.P. Limongelli (Politecnico Milan), E. Chatzi (ETH Zürich), G. Lombaert and E. Reynders (both KU Leuven).

After a brief review of vibration based damage identification methods, three different algorithms for damage identification are applied to the case of the benchmark Z24 bridge. Data-driven as well as model-based methods are discussed, including input-output algorithms for taking into account the effect of environmental and/or operational sources on the variability of damage features. A further class of data-driven methods that use finite element information is finally introduced as a possible future development. [35]

7.5.6. Structural system reliability updating with subspace-based damage detection information

Participant: Michael Doehler.

This work is in collaboration with S. Thöns (DTU).

Damage detection systems and algorithms (DDS and DDA) provide information of the structural system integrity in contrast to e.g. local information by inspections or non-destructive testing techniques. However, the potential of utilizing DDS information for the structural integrity assessment and prognosis is hardly exploited nor treated in scientific literature up to now. In order to utilize the information provided by DDS for the structural performance, usually high computational efforts for the pre-determination of DDS reliability are required. In this work, an approach for the DDS performance modelling is introduced building upon the non-destructive testing reliability which applies to structural systems and DDS containing a strategy to overcome the high computational efforts for the pre-determination of the DDS reliability. This approach takes basis in the subspace-based damage detection method and builds upon mathematical properties of the damage detection algorithm. Computational efficiency is gained by calculating the probability of damage indication directly without necessitating a pre-determination for all damage states. The developed approach is applied to a static, dynamic, deterioration and reliability structural system model, demonstrating the potentials for utilizing DDS for risk reduction. [30]

7.5.7. Structural system model updating based on different sensor types

Participants: Dominique Siegert, Xavier Chapeleau, Ivan Guéguen.

Detecting and quantifying early structural damages using deterministic and probabilistic model updating techniques can be achieved by local information in a form of optical strain measurement. The strategy consists in updating physical parameters associated to damages, such as Young's modulus, in order to minimize the gap between the numerical strain obtained from finite element solves and the strain sensor outputs. Generally, the damage estimation is an ill-posed inverse problem, and hence requires regularization. Herein, three model updating techniques are considered involving different type of regularization: classical Tikhonov regularization, constitutive relation error based updating method and Bayesian approach [21]. This work follows an experimental campaign carried out on a post tensioned concrete beam with the aim of investigating the possibility to detect early warning signs of deterioration based on static and/or dynamic tests. Responses of a beam were measured by an extensive set of instruments consisting of accelerometers, inclinometers, displacement transducers, strain gauges and optical fibers. [18].

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Contracts with Industry

8.1.1. PhD project with EDF – Electrical device ageing monitoring

Participants: Nassif Berrabah, Qinghua Zhang.

A joint PhD project between Inria and EDF (ElectricitÈ de France) has been started since December 2014. The purpose of this study is to develop methods for the monitoring of electrical instruments in power stations, in order to prevent failures caused by ageing or accidental events. This project is funded by EDF and by the ANRT agency for three years.

8.1.2. Contracts with SVS

Participants: Laurent Mevel, Michael Doehler.

I4S is doing technology transfer towards SVS to implement I4S technologies into ARTEMIS Extractor Pro. This is done under a royalty agreement between Inria and SVS .

In 2014, the damage detection toolbox has been launched http://www.svibs.com/products/ARTeMIS_Modal_Features/Damage_Detection.aspx.

In 2015, SVS and Inria have earned an Innobooster grant to help transfer algorithms in 2016 Artemis Extractor Pro.

In 2016, uncertainty quantification for modal analysis has been launched http://www.svibs.com/newsletter/newsletter_2016_09.aspx.

8.1.3. Contracts with A3IP

Participant: Vincent Le Cam.

Since 2008, IFSTTAR has licensed the company A3IP to sell licenses of the PEGASE 1 platform (previous version of PEGASE 2 as mentioned above). A3IP sells them to to companies, laboratories or any third-party partner interested in in-situ monitoring (SHM) with smart and wireless sensors. Since 2008, about 1000 of PEGASE 1 units have been sold, plus hundreds of the following items:

- daughter boards: 3G / Ethernet communications, Analog to Digital data acquisition...
- sensors: accelerometer, strain gauges, temperature...
- specific packaging to make the PEGASE 1 solution ready to use in waterproof conditioning

For example, in 2016, A3IP has provided a complete panoply of PEGASE-1 Vibration Monitoring system with more than 30 PEGASE1 units to ensure the monitoring of the new High Speed Train line in west of France (Bretagne Pays de la Loire high speed railway).

This non exclusive license is clearly a success in terms of dissemination.

8.1.4. Contract with SNCF: DEMETER

Participant: Vincent Le Cam.

DEMETER is one of the major projects for I4S in terms of strategy, scientific and technological impact.

DEMETER is a meta project whose global objective is the validation of the contribution of the Internet of Things (IOT) applied to the Health Monitoring of Railways Items. SNCF and IFSTTAR have signed a roadmap for safety relevant items, where wireless monitoring and smart algorithms could bring strong improvements to SNCF in terms of real-time maintenance or predictive maintenance. Those items are, amongst others:

- Crossing engine motor monitoring
- Needle motor monitoring
- Axel counter monitoring
- Train detection pedal monitoring

In each case, a prototype of a specific PEGASE 2 sensor is designed, installed along in-situ railways lines under exploitation and data are transmitted wireless to the cloud supervisor at IFSTTAR for evaluation in SHM algorithms. IFSTTAR's engineers Arthur Bouche, Laurent Lemarchand and David Pallier are contributing to this project.

In particular, SNCF and IFSTTAR are able to perform the entire validation process quickly in few months: from the algorithm to the electronic design and installation. In 2016, the consortium reached 2 milestones: the PEDAL-LORA monitoring sensor has been awarded the European Railway Cluster Price in railway innovation; this system is now becoming an industrial product, directly designed by a third-party company for SNCF. In 2017, the roadmap will be extended with a specific focus on SHM algorithms implementation to help SNCF moving from big data to smart data.

8.1.5. Contracts with SDEL-CC (VINCI Group)

Participant: Vincent Le Cam.

In 2016, a contract has been signed with the company SDEL-CC, 100% daughter of the VINCI Group, Energy department. The project exploits the unique time stamp capacity of PEGASE 2 up to 50 nanosecond, independently of distances in the network of PEGASE2 nodes. The synchronization capacity is employed to design a sensor prototype based on PEGASE2 to time-stamp the current wave after a lightning impact on a high-voltage line. By knowing the exact time, the wave can be seen at each extremity of the electrical line to localize accurately the lightning impact point. IFSTTAR's engineers Arthur Bouche and Laurent Lemarchand have contributed to this project.

During 2016, we have improved its embedded algorithms on PEGASE 2 platform to:

- take into account some specific GPS frames that output from its GPS receiver and give practical information on time drift
- take into account the temperature effect
- auto compute the real quartz period on each specific PEGASE 2 board

Two PEGASE 2 platforms are now able to time stamp an event with an accuracy of less that 10 nanoseconds. This leads to a precision of around 3 m for Lightning localization.

In 2017 in situ validation will be achieved on a real operated electric line.

8.1.6. Collaboration with SIEMENS : CityVal Rennes

Participant: Jean Dumoulin.

A first Winter season measurements campaign on the 100m metro structure mock-up built at IFSTTAR test track facilities in Nantes was carried out in 2016. It was completed by in situ instrumentation including coupling of infrared thermography with other measurements techniques for long term monitoring during several months. A new campaign is under preparation and will be launched in 2017. This collaboration is also connected with the new automated metro line B under construction in Rennes.

9. Partnerships and Cooperations

9.1. Regional Initiatives

9.1.1. MONEOL - project with CEAtech Pays de Loire

Participants: Ivan Guéguen, Guillaume Gautier, Laurent Mevel.

Type: CEAtech PDL

Objectif: Modal analysis of wind turbines using new sensors

Duration: 11/2015 to 11/2017.

Coordinator: Louis Marie Cotineau (IFSTTAR)

Inria contact: Guillaume Gautier

Abstract: The MONEOL project aims to demonstrate the feasibility of using Morphosense as a vibration monitoring system for wind turbines. It is proposed to set up a demonstrator consisting of a monitoring system placed in the mast of the wind turbine, a vibration analysis system and a visualization of the vibratory state at the CEA-Tech premises, located on the Technocampus Ocean of Nantes allowing to visualize in real time (quasi) the modal deformations of the mast of the wind turbine. This system consists of the following elements:

The demonstrator consists of the monitoring system placed in the wind turbine of a video screen displaying in real time indicators to evaluate the state of health of the structure:

- Modal parameters (eigen frequencies, modal damping, modal deformations) over time and associated uncertainties.
- Indicators of detection and localization of damage.

The demonstrator will also be able to display a video of the wind turbine in operation. In order to validate the Morphosense sensor, a reference system is added to it, consisting of conventional accelerometer sensors.

9.1.2. Interactive Communication (InterCom): Massive random access to subsets of compressed correlated data

Participants: Jean Dumoulin, Antoine Crinière.

Type: Labex COMINLABS

Objectif: Massive random access to large-scale sensor network (Smart Cities)

Duration: Since November 2016 to Nov. 2019.

Coordinator :Aline Roumy, Thomas Maugey (Sirocco), Jean Dumoulin (I4S)

Partners: Elsa Dupraz (Lab-STICC), Aline Roumy (IRISA, Sirocco team), Michel Kieffer (L2S), Thomas Maugey(IRISA, Sirocco team), CentraleSupelec, Univ. Paris Sud.

Inria contact: Jean Dumoulin

Abstract: This project aims to develop novel compression techniques allowing massive random access to large databases. Indeed, we consider a database that is so large that, to be stored on a single server, the data have to be compressed efficiently, meaning that the redundancy/correlation between the data have to be exploited. The dataset is then stored on a server and made available to users that may want to access only a subset of the data. Such a request for a subset of the data is indeed random, since the choice of the subset is user-dependent. Finally, massive requests are made, meaning that, upon request, the server can only perform low complexity operations (such as bit extraction but no decompression/compression).

Algorithms for two emerging applications of this problem will be developed: Free-viewpoint Television (FTV) and massive requests to a database collecting data from a large-scale sensor network (such as Smart Cities) in which I4S is involved.

9.1.3. MAG2C-Pont Tabarly

Participants: Ivan Guéguen, Jean Dumoulin.

Type: GIS

Objectif: bridge instrumentation

Duration: Since 2014

Coordinator: LIRGEC

Partners: IFSTTAR, CSTB, Nantes Métropole, Université de Nantes

Inria contact: Ivan Guéguen

Abstract: The project deals with the instrumentation of the Tabarly Bridge.

Based on accelerometer measurements, the vibration behaviour will be monitored and structural defects detected. Coupled with a wireless data transmission system type or wifi 3g, remote monitoring is envisaged. The different objectives are

- Experimentation on a bridge
- Equipment qualification in real conditions over long term
- Apply different vibration processing algorithms
- Monitoring and detection
- Measurement database

An accelerometer-based distributed network on the structure is installed and connected to a data acquisition system and a modem 3g for continuous remote measurements, which will be available on the internet.

9.1.4. MAG2C-MOSIWIND (MOnitoring of Structural Integrity of an onshore WIND turbine's slab foundation and tower)

Participants: Xavier Chapeleau, Ivan Guéguen.

Type: GIS

Objectif: MOnitoring of Structural Integrity of an onshore WIND turbine's slab foundation and tower

Duration: Since 2015

Coordinator : LIRGEC

Partners: IFSTTAR, CSTB, Nantes Métropole, Université de Nantes, ECN, Valorem, Valréa and Valémo

Inria contact: Xavier Chapeleau

Abstract: The project deals with the instrumentation of an onshore WIND turbine's slab foundation and tower. The aim is to experiment sensors and methods for structural integrity monitoring of an onshore wind turbine under real conditions and to qualify them over long term. Before casting, the concrete slab foundation (20m in diameter, 3.85m high, 450m3 of concrete, 48T of reinforcement) was first instrumented with continuous optical fibers, optical strain gauges, temperature sensors and accelerometers. Afterwards, accelerometers were placed in the mast. Data obtained by these different sensors will help, on the one hand, to monitor changes in the dynamic behavior of the structure in order to verify that they remain within the limits fixed during the design and, on the other hand, to detect any damage that could be critical for the safety of the structure. For this, SSI methods under ambient vibration will be applied.

9.1.5. Collaboration with GEM

Participants: Laurent Mevel, Michael Doehler, Md Delwar Hossain Bhuyan.

Md Delwar Hossain Bhuyan has started a PhD on Damage localisation on offshore platforms, The thesis is co-directed by L. Mevel and F. Schoefs from GEM, Nantes, with supervision shared with M. Doehler and Y. Lecieux from GEM. It is funded by the Brittany region for 3 years.

9.2. National Initiatives

9.2.1. High speed rail track instrumentation

Participant: Ivan Guéguen.

Type: IRT

Objective: rail track SHM

Duration: 11/2014 to 11/2018

Coordinator: RAILENIUM

Partners: IFSTTAR, EIFFAGE, RFF, LGCgE

Inria contact: Ivan Guéguen

Abstract: This project aims at instrumenting multiple sections of a high-speed route (classical section with granular layer, transition zone). The proposed instrumentation concerns all the different layers of the structure, and is designed to allow monitoring of the overall track behavior.

The instrumentation will include:

- A weather station for environmental conditions (temperature, precipitation on the site).

- Accelerometers, to monitor the dynamic behavior of the track, with measurements at several levels: the hammer beams on top of the grave-bitumen layer, on top of the soil.
- Instrumentation of severe bitumen strain gauges for measuring the longitudinal and transverse tensile strains, and temperature probes (top and bottom layer). This instrumentation will estimate the fatigue life of the GB, temperature changes in this layer, and will calculate a temperature equivalent to the layer of GB.
- Instrumentation subgrade by means of measurement gauges at the top of the vertical deformation of the soil, and TDR probes to measure changes in water content. Its objective is to measure the levels of distortion in the upper part of the soil, and their variations, in conjunction with the seasonal variations in water content.
- An anchored sensor, measuring the total deflection between the top of the GB and a reference point that is 4 m deep. This sensor will measure the total displacement of the structure beneath the ballast (GB + layer of granular soil leveling + support). These will also serve as a reference for comparison with the movements deducted from accelerometer measurements.
- Continuous optical fiber, to measure static permanent deformation in the transverse direction over the entire width of the structure at the base of the sub-layer.

9.2.2. ANR Resbati

Participant: Jean Dumoulin.

Type: ANR

Objectif: In-situ measurements of thermal wall resistance

Duration: 10/2016 to 10/2019

Coordinator: Laurent ibos

Partners : IFSTTAR, CERTES, CEREMA, CSTB, LNE, THEMACS, AFNOR

Inria contact: Jean Dumoulin

Abstract: Thermal insulation of opaque walls remains an essential point for improving the energy efficiency in buildings. Indeed, the number of badly insulated buildings in France is still very important. In addition, current thermal regulations set high requirements in terms of thermal insulation and will continue to be more rigorous as new building will be energy-positive with the French RT2020. However, there is no systematic method for measuring the thermal insulation level of the building walls. Their thermal performance must be controlled for renovation of the building, during its construction, for its delivery or during use. The need of a method of in-situ control of walls is more relevant than ever. Such a measurement at the wall level is an interesting complement to global methods (co-heating, etc.) that concern the whole building energy balance. The physical parameter representing the quality of the wall thermal insulation is its thermal resistance. Currently, methods for measuring this parameter exist, either in the form of laboratory or exploratory methods, or in the form of international standards or draft standards. However, each of these methods does not meet all the conditions guaranteeing a general measurement: use on any type of wall and at any time of the year, low measurement duration, ease of use, moderate cost. The RESBATI project (in-situ measurement of the thermal resistance of building walls) aims at developing an in-situ measurement device that respects these specifications. The measuring means is infrared thermography in active approach. The uncertainty and the limitations of the measurement will be identified during the project. Infrared thermography in passive mode has demonstrated for many years its ability to reveal the presence of insulation defects in buildings. However, it is essentially a qualitative tool. The active approach of infrared thermography is not very used for building investigation and is a promising way for obtaining quantitative information such as the thermal resistance of the wall to investigate. Indeed research results have already shown that this approach could be used to obtain quantitative estimations of the thermal resistance of opaque building walls. The RESBATI project will demonstrate the potential of the active approach so that control can be performed in any season, for any type of building and any use (occupied or not) and quickly. The passive approach might nevertheless be used as a complement because it does not require the use of additional equipment ensuring the thermal load of the wall to diagnose and provides access to larger wall surfaces to analyze. The consortium brings complementary partners together working at different levels of the building: research laboratories, technical center, national metrology laboratory, company and standards organization. The advanced knowledge and past achievements of the various partners on the subject make it possible to develop such a method with measurement uncertainty and the associated prototypes. Many facilities will be available for qualification of prototypes: climate rooms for laboratory testing, existing buildings for in-situ qualifications. Thus, a wide variety of walls (structure and isolation level) can be tested. Moreover, these buildings have different uses (residential or service buildings). In conclusion of the project, measurements will be carried out by future end-users of the device.

9.2.3. Equipex Sense-City

Participants: Jean Dumoulin, Laurent Mevel, Antoine Crinière.

Through the ADT Cloud2SM, participation of I4S in SenseCity was possible. IFSTTAR's SensorBox developed by Jean Dumoulin was installed and presented at SEnseCity Kick off and is installed on-site. Cloud2IR and Cloud2SM software have been deployed within the ADT of A. Crinière. (http://sense-city.ifsttar.fr/)

9.3. European Initiatives

9.3.1. FP7 & H2020 Projects

9.3.1.1. Built to Specifications (Built2Spec)

Participants: Jean Dumoulin, Alexandre Nassiopoulos, Jordan Brouns.

Type: Horizon 2020

Defi: Model Driven Physical Systems Operation

Objectif: Reduce the gap between a building?s designed and as-built energy performance.

Duration: January 2015 to January 2019

Coordinator: Manager and project head : NOBATEK, Germain Adell. For CERMA : Marjorie Musy

Inria teams I4S

Inria contact: J. Dumoulin

Partners: Consortium of 20 Public and Industrial actors

Website: http://built2spec-project.eu/

Abstract: Built to Specifications (Built2Spec) is a Horizon 2020 EU-funded project involving 20 European partners that seeks to reduce the gap between a building?s designed and as-built energy performance. To do this, the project will put a new set of breakthrough technological advances for self-inspection checks and quality assurance measures into the hands of construction professionals. This collection of smart tools will help building stakeholders at all levels in meeting EU energy efficiency targets, new build standards and related policy goals.
Built2Spec will deliver a new set of tools:

- 3D and Imagery Tools
- Building Information Modelling (BIM)
- Smart Building Components
- Energy Efficiency Quality Checks
- Indoor Air Quality Tools
- Airtightness Test Tools with Air-pulse Checks

- Thermal Imaging Tools
- Acoustic Tools

All connected to a Virtual Construction Management Platform supporting the collection and sharing of all project data, from initial design to the delivery. During the project, this platform will be integrated into the operations of small and medium-sized enterprise (SME) contractors, large construction firms and end user clients directly within the consortium and work program activities, assuring systematic and scientific performance measures, feedback and powerful exploitation.

9.3.1.2. INFRASTAR (Innovation and Networking for Fatigue and Reliability Analysis of Structures – Training for Assessment of Risk)

Participant: Xavier Chapeleau.

Call: H2020-MSCA-ITN-2015 (Horizon 2020 – Marie-Sklodowska Curie Actions – Innovative Training Networks)

Type of Action: MSCA-ITN-ETN

Objectif: Reduce the gap between a building?s designed and as-built energy performance.

Duration: 48 months since 2016 May 1st

Coordinator: Odile Abraham (IFSTTAR)

Academic and industrial Partners: IFSTTAR, UNIVERSITY OF AALBORG, BAM, EPFL, GuD Consult Gmbh, COWI A/S, NeoStrain, PHIMECA

Inria contact: X. Chapeleau

Website: http://infrastar.eu/

Abstract: The aim of INFRASTAR project is to develop tools combining modeling and measurements for the prediction of the fatigue behavior of concrete structures (bridges and foundations of wind turbines) with the ultimate objective of establishing an efficient strategy for inspection and reinforcement operations. In the second half of 2016, 12 young researchers were recruited to carry out and cross-examine research on monitoring and auscultation (WP 1), structural models (WP 2) and reliability of approaches for decision-making (WP 3). In this project, a phd student (Antoine Bassil) was recruited (Nov. 2016) on the fatigue monitoring of concrete structure by fibre-optic sensors.

9.3.2. Collaborations in European Programs, Except FP7 & H2020

9.3.2.1. European Research Network on System Identification (ERNSI)

Participants: Qinghua Zhang, Michael Doehler, Laurent Mevel.

The I4S project-team is involved in the activities of the European Research Network on System Identification (ERNSI) federating major European research teams on system identification. Modeling of dynamical systems is fundamental in almost all disciplines of science and engineering, ranging from life science to process control. System identification concerns the construction, estimation and validation of mathematical models of dynamical physical or engineering phenomena from experimental data.

9.3.2.2. COST Action TU 1402

Participants: Michael Doehler, Laurent Mevel.

L. Mevel is member of the management committee of the COST Action.

M. Doehler is co-leader of working group 2 “SHM strategies and structural performance” and member of the steering committee.

Type: COST

Objectif: Quantifying the value of structural health monitoring

Duration: 11/2014 - 11/2018

Coordinator: S. Thoens (DTU Denmark)

Partner: 23 countries, see http://www.cost.eu/COST_Actions/tud/Actions/TU1402

Inria contact: Laurent Mevel

Abstract: This COST Action enhances the benefit of Structural Health Monitoring (SHM) by novel utilization of applied decision analysis on how to assess the value of SHM - even before it is implemented. This improves decision basis for design, operation and life-cycle integrity management of structures and facilitates more cost efficient, reliable and safe strategies for maintaining and developing the built environment to the benefit of society. SHM is increasingly applied for collecting information on loads and aggressive environments acting on structures, structural performances, deterioration processes and changes in the use of structures. However, there is an urgent need to establish a better understanding of the value of SHM before its implementation, together with practically applicable methods and tools for its quantification. This Action thus aims to develop and describe a theoretical framework, together with methods, tools, guidelines, examples and educational activities, for the quantification of the value of SHM. The COST Action will be conducted with the support of the Joint Committee on Structural Safety (JCSS). The networks of researchers and industries established during COST Actions TU0601, C26, E55 and E24, the EU FP7 project IRIS, the Marie Curie Network SmartEn and the JCSS will ensure visibility, impact and dissemination.

9.3.3. Other European Programs

9.3.3.1. Innobooster

Participants: Michael Doehler, Laurent Mevel.

Together with SVS, we got the Danish Innobooster innovation grant “Robust Operational Modal Analysis using Modal Uncertainty Quantification” 2015-2016, for industrial research and transfer. The result of the development in this project is the transfer of our uncertainty quantification algorithm [19] to SVS’ ARTeMIS software http://www.svibs.com/newsletter/newsletter_2016_09.aspx.

9.4. International Initiatives

9.4.1. Informal International Partners

9.4.1.1. Collaboration with CNR, Italy

Participants: Jean Dumoulin, Nicolas Le Touz.

Non destructive testing on outdoor structures by coupling infrared thermography with ground penetrating radar is one of the topic addressed in this collaboration. A new one about TerHertz is starting.

9.4.1.2. Collaboration with British Columbia University, Canada

Participants: Laurent Mevel, Michael Doehler, Saeid Allahdadian.

Saeid Allahdadian is currently PhD student of professor Carlos Ventura in Vancouver. Following our recent papers, Michael Doehler has been invited to co-supervise the PhD of Saeid Allahdadian starting in 2015 for 3 years.

9.4.1.3. Collaboration with BAM, Germany

Participants: Laurent Mevel, Michael Doehler, Eva Viefhues.

Eva Viefhues is currently PhD student of Laurent Mevel and Michel Doehler in Berlin, financed by BAM. M. Doehler is also associate researcher of the BAM institut since 2016.

9.4.1.4. Collaboration with Politecnico di Milano, Italy

Participants: Michael Doehler, Dominique Siegert, Ivan Guéguen, Xavier Chapeleau.

During COST Action TU 1402 and M.P. Limongelli's research stay at IFSTTAR, collaboration with Politecnico di Milano has started, resulting in several joint publications in 2016 [35], [18], [21]. A joint Master student project is in progress, and a french-italian PhD project is planned.

9.4.2. Participation in Other International Programs

The team has been awarded a MITACS grant. It allowed us to host S. Allahdadian for 3 months in 2016.

9.5. International Research Visitors

9.5.1. Visits of International Scientists

S. Allahdadian from British Columbia University has visited us for 3 months in 2016 thanks to a MITACS grant.

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events Selection

10.1.1.1. Member of the Conference Program Committees

J.Dumoulin is

- member of the scientific committee of the GI Division (Geosciences Instrumentation and Data Systems) of EGU (European Geosciences Union) for infrastructure instrumentation and monitoring since April 2013. (http://www.egu.eu/gi/structure/)
- member of the scientific committee of QIRT (quantitative Infrared Thermography) since February 2014 (http://www.qirt.org/)
- organizer and chair of a session at EGU 2016 (http://www.egu2016.eu/).

L. Mevel

- is member of the EWSHM scientific committee.
- is member of the IOMAC scientific committee.

V. Le Cam is head and general secretary of the EWSHM scientific committee.

Q. Zhang:

- Member of IFAC Technical Committee on Modelling, Identification and Signal Processing.
- Member of IFAC Technical Committee on Fault Detection, Supervision and Safety of Technical Processes.

M. Doehler

- was organizer of an invited session at EWSHM 2016.
- session organizer at two COST workshops (http://www.cost-tu1402.eu/events).

10.1.1.2. Reviewer

V. Le Cam was session chairman for the EWSHM 2016 in Bilbao

L. Mevel was reviewer and session chairman for the EWSHM 2016 in Bilbao, and reviewer for IFAC WC 2017

Q. Zhang was reviewer for CDC 2016, IFAC WC 2017

M. Doehler was reviewer and session chairman for EWSHM 2016 in Bilbao, and reviewer for MED 2016, IFAC WC 2017

J. Dumoulin was reviewer and session chairman for QIRT 2016 and at EGU 2016 in GI division

10.1.2. Journal

10.1.2.1. member of the Editorial Boards

L. Mevel is member of the editorial board of journal of Mathematical Problems in Engineering.

L. Mevel is member of the editorial board of journal of Shock and Vibration.

Q. Zhang is member of the editorial board of the journal of Intelligent Industrial Systems.

J. Dumoulin is member of the editorial board of the journal of Quantitative Infrared Thermography.

J. Dumoulin is member of the editorial board of the journal of Geoscientific Instrumentation and Data Systems.

10.1.2.2. Reviewer - Reviewing activities

X. Chapeleau was reviewer for Journal: Sensors and Journal of Civil Structural Health Monitoring

L. Mevel was reviewer for Mechanical Systems and Signal Processing, journal of Sound And Vibration and Journal of Control and SHM.

M. Doehler was reviewer for Automatica, International Journal of Control, International Journal of Systems Science, Mechanical Systems and Signal Processing, Journal of Sound and Vibration, Mathematical Problems in Engineering, Smart Materials and Structures, Journal of Intelligent Material Systems and Structures

J. Dumoulin was reviewer for IEEE Transactions on Instrumentation and Measurement, Quantitative Infrared Thermography Journal, Optics and Lasers in Engineering journal , Journal Cultural Heritage, International Journal of Architectural Heritage, Journal of Geophysics and Engineering, Research in Nondestructive Evaluation

10.1.3. Invited Talks

J. Dumoulin was invited speaker at the 4th Youth in Conservation of Cultural heritage YOCOCU 2016, 21-23 September 2016, Madrid, Spain.

J. Dumoulin was invited keynote speaker at ERICE in October 2016.

J. Dumoulin was invited speaker at SFT 2016 in March 2016.

M. Mogoro (from SNCF) and V. Le Cam have been invited to give a keynote at the first SHM Conference dedicated to Railway in Qingdao, China, on October 2016 http://www.crrcgc.cc/iwshm-rs/english/.

V. Le Cam have was invited speaker at the "NDT Conference organized by Airbus Group at Pondichery, India, January 2016"

10.1.4. Scientific Expertise

10.1.4.1. Method and device for localizing faults in an electrical cable

Participants: Nassif Berrabah, Qinghua Zhang.

In modern engineering systems, fault diagnosis is frequently an integrated functionality for various components, but rarely for electrical cables. The fast development of electronic devices is accompanied by more and more connecting cables. The reliability of electrical connections becomes a crucial issue because of their large number. Moreover, some cables are operated under severe conditions, such as extreme temperature, nuclear radiation, humidity, mechanical strain, etc.. Based on the work reported in Section 7.5.1, a patent has been registered at INPI jointly by EDF and Inria [46]. It is about a method for detecting, localizing and quantifying resistive faults in a cable, by means of estimating the series resistance per unit length distributed along the cable from reflectometry measurements made at the ends of the cable. Its fast numerical computation makes it suitable for real time applications.

10.1.4.2. Scientific Expertise in European Calls

Participant: Vincent Le Cam.

V. Le Cam : Expertise of a specific EUROPEAN SME project in the call EUROSTARS.

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

Licence Professionnelle TAM : J. Dumoulin, thermographie infrarouge active, 16h, Université Paris-Est, France

Master 2 MMMRI, (Maintenance et Maîtrise des Risques Industriels) , J. Dumoulin, contrôle non destructif par thermographie infrarouge active, 12h, Université Paris-Est, France

Master 2 ITII, J. Dumoulin, , BTP, module Maintenance et réhabilitation des ouvrages, « Transferts thermiques dans les Structures : Des principes physiques à l'application sur site réel », 12 h, Ecole Centrale de Nantes(ECN).

Master Système Communicant Mobile, V. Le Cam, embedded systems under Linux Operating System, 12h, Polytech Nantes, France

Master Civil engineering, V. Le Cam, Structural Monitoring, 4h, Université de Nantes, France

Licence 3 SEICOM, V. Le Cam, 3h, SHM and smart grids, Université de Nantes, France

Licence 3 SEICOM, V. Le Cam, 8h, TP, SHM and smart grids, Université de Nantes, France

ESEO, V. Le Cam, 16h, TP, embedded systems under Linux Operating System, France

Polytech, V. Le Cam, 14h, TP, embedded systems under Linux Operating System, France

Master 1 informatique, M. Doehler, 24 TD projet recherche, Université de Rennes 1 & ENS Rennes, France

Licence Pro Mesures physiques, X. Chapeleau, Mesures optiques, 15h, IUT de St Nazaire, Université de Nantes, France

10.2.2. Supervision

PhD : Antoine Bassil, *Fibre-optic sensor for fatigue monitoring*, D. Leduc, O. Abraham and X. Chapeleau, Ecole doctorale SPIGA, Université de Nantes, since November 2016.

PhD : Delwar Hossain Bhuyan, *Damage localisation on offshore platforms*, L. Mevel and M. Doehler, Ecole doctorale MATISSE, Université de Rennes 1, since November 2014

Guillaume Gautier's post-doctoral project on morphosense system monitoring, L. Mevel, 2015-2017.

PhD : Nassif Berrabah, Electrical cable ageing monitoring , Q. Zhang, Ecole doctorale MATISSE, Université de Rennes 1, since November 2014

PhD : Nicolas Le Touz. *Design and study of positive energy transport infrastructures: from thermomechanical modeling to the optimization of such energy systems* J. Dumoulin. at Ecole Centrale Nantes (ECN) since 2015.

PhD : Thibault Toullier. *Simultaneous characterization of the radiative properties and temperatures of envelopes of structures in natural environment by multispectral infrared thermography* L. Mevel, J. Dumoulin and M. Doehler. Ecole doctorale MATISSE, Université de Rennes 1, since November 2016.

PhD : Saeid Allahdadian, *Methods for vibration-based damage assessment*, M. Doehler, University of British Columbia, Canada, since 2015.

PhD : Eva Viefhues, *Statistical damage localization for civil structures*, L. Mevel and M. Doehler, Ecole doctorale MATISSE, Université de Rennes 1, since November 2016.

J. Dumoulin is associate professor at Laval University, Canada.

M. Doehler is associate researcher at BAM, Germany.

10.3. Popularization

J. Dumoulin was in charge of hybrid/solar road that was demonstrated at COP21 and COP22.

11. Bibliography

Major publications by the team in recent years

[1] J. BROUNS, A. CRINIÈRE, J. DUMOULIN, A. NASSIOPOULOS, F. BOURQUIN.*Diagnostic de structures de Génie Civil : Identification des propriétés spatiales et de la surface d'un défaut*, in "SFT 2014", Lyon, France, Société Française de Thermique, May 2014, https://hal.inria.fr/hal-01082184.

[2] A. CRINIÈRE, J. DUMOULIN, C. IBARRA-CASTANEDO, X. MALDAGUE.*Inverse model for defect characterisation of externally glued CFRP on reinforced concrete structures: comparative study of square pulsed and pulsed thermography*, in "Quantitative InfraRed Thermography Journal", March 2014, vol. 11, n° 1, p. 84-114 [*DOI :* 10.1080/17686733.2014.897512], https://hal.archives-ouvertes.fr/hal-01081174.

[3] J. DUMOULIN, V. BOUCHER.*Infrared thermography system for transport infrastructures survey with inline local atmospheric parameter measurements and offline model for radiation attenuation evaluations*, in "Journal of Applied Remote Sensing", 2014, vol. 8, n° 1, p. 084978–084978.

[4] J. DUMOULIN, A. CRINIÈRE, R. AVERTY.*The detection and thermal characterization of the inner structure of the 'Musmeci' bridge deck by infrared thermography monitoring*, in "Journal of Geophysics and Engineering", December 2013, vol. 10, n° 6, 17 [*DOI :* 10.1088/1742-2132/10/6/064003], https://hal.inria.fr/hal-01081320.

[5] M. DÖHLER, L. MEVEL.*Fast Multi-Order Computation of System Matrices in Subspace-Based System Identification*, in "Control Engineering Practice", September 2012, vol. 20, n° 9, p. 882–894.

[6] M. DÖHLER, L. MEVEL.*Modular Subspace-Based System Identification from Multi-Setup Measurements*, in "IEEE Transactions on Automatic Control", November 2012, vol. 57, n° 11, p. 2951–2956.

[7] M. DÖHLER, L. MEVEL.*Efficient Multi-Order Uncertainty Computation for Stochastic Subspace Identification*, in "Mechanical Systems and Signal Processing", June 2013, vol. 38, n° 2, p. 346–366.

[8] M. DÖHLER, L. MEVEL.*Subspace-based fault detection robust to changes in the noise covariances*, in "Automatica", September 2013, vol. 49, n° 9, p. 2734–2743 [*DOI :* 10.1016/J.AUTOMATICA.2013.06.019], https://hal.inria.fr/hal-00907662.

[9] A. JHINAOUI, L. MEVEL, J. MORLIER.*A new SSI algorithm for LPTV systems: application to a hinged-bladed helicopter*, in "Mechanical Systems and Signal Processing", January 2014, vol. 42, n° 1, p. 152–166.

[10] P. LAIR, J. DUMOULIN, P. MILLAN.*Inverse method for flux characterization using infrared thermography in die forging*, in "Numerical Heat Transfer, Part A Applications", 1998, vol. 33, n° 3, p. 267–277.

[11] F. LOETE, Q. ZHANG, M. SORINE.*Experimental validation of the inverse scattering method for distributed characteristic impedance estimation*, in "IEEE Transactions on Antennas and Propagation", 2015, vol. 63, n° 6, 7 [*DOI :* 10.1109/TAP.2015.2417215], https://hal.inria.fr/hal-01231807.

[12] L. MARIN, M. DÖHLER, D. BERNAL, L. MEVEL.*Robust statistical damage localization with stochastic load vectors*, in "Structural Control and Health Monitoring", March 2015, vol. 22, n° 3.

[13] M. ZGHAL, L. MEVEL, P. DEL MORAL.*Modal parameter estimation using interacting Kalman filter*, in "Mechanical Systems and Signal Processing", August 2014, vol. 47, n^{o} 1, p. 139–150.

[14] Q. ZHANG, M. SORINE, M. ADMANE.*Inverse Scattering for Soft Fault Diagnosis in Electric Transmission Lines*, in "IEEE Transactions on Antennas and Propagation", 2011, vol. 59, n^{o} 1, p. 141 - 148, https://hal.inria.fr/inria-00365991.

Publications of the year

Articles in International Peer-Reviewed Journal

[15] J. BROUNS, A. NASSIOPOULOS, F. BOURQUIN, K. LIMAM.*Dynamic building performance assessment using calibrated simulation*, in "Energy and Buildings", 2016, vol. 122, 15 [*DOI :* 10.1016/J.ENBUILD.2016.04.015], https://hal.inria.fr/hal-01354107.

[16] J. BROUNS, A. NASSIOPOULOS, K. LIMAM, F. BOURQUIN.*Heat source discrimination in buildings to reconstruct internal gains from temperature measurements*, in "Energy and Buildings", January 2017, vol. 135, p. pp 253-262 [*DOI :* 10.1016/J.ENBUILD.2016.11.041], https://hal.archives-ouvertes.fr/hal-01409762.

[17] M. DÖHLER, L. MEVEL, Q. ZHANG.*Fault Detection, Isolation and Quantification from Gaussian Residuals with Application to Structural Damage Diagnosis*, in "Annual Reviews in Control", 2016, vol. 42, p. 244-256 [*DOI :* 10.1016/J.ARCONTROL.2016.08.002], https://hal.inria.fr/hal-01376804.

[18] M. P. LIMONGELLI, D. SIEGERT, E. MERLIOT, J. WAEYTENS, F. BOURQUIN, R. VIDAL, V. LE CORVEC, I. GUEGUEN, L.-M. COTTINEAU.*Damage detection in a post tensioned concrete beam – Experimental investigation*, in "Engineering Structures", September 2016, vol. 128, p. 15 - 25 [*DOI :* 10.1016/J.ENGSTRUCT.2016.09.017], https://hal.inria.fr/hal-01400881.

[19] P. MELLINGER, M. DÖHLER, L. MEVEL.*Variance estimation of modal parameters from output-only and input/output subspace-based system identification*, in "Journal of Sound and Vibration", June 2016, vol. 379, p. 1-27 [*DOI :* 10.1016/J.JSV.2016.05.037], https://hal.inria.fr/hal-01328435.

[20] B. NI, Q. ZHANG.*Stability of the Kalman filter for continuous time output error systems*, in "Systems and Control Letters", 2016, vol. 94, p. 172 - 180 [*DOI :* 10.1016/J.SYSCONLE.2016.06.006], https://hal.inria.fr/hal-01376819.

[21] J. WAEYTENS, B. ROSIĆ, P.-E. CHARBONNEL, E. MERLIOT, D. SIEGERT, X. CHAPELEAU, R. VIDAL, V. LE CORVEC, L.-M. COTTINEAU.*Model updating techniques for damage detection in concrete beam using optical fiber strain measurement device*, in "Engineering Structures", 2016 [*DOI :* 10.1016/J.ENGSTRUCT.2016.08.004], https://hal.inria.fr/hal-01379214.

Invited Conferences

[22] J. DUMOULIN.*Considerations on outdoor application of Infrared Thermography for transport infrastructures survey*, in "Fourth Mediterranean International Workshop on Photoacoustic & Photothermal Phenomena", Erice, Italy, October 2016, https://hal.inria.fr/hal-01377102.

[23] J. DUMOULIN.*Review of non destructive testing with infrared thermography in cultural heritage: from indoor to outdoor practices: Keynote by Jean Dumoulin*, in "YOCOCU 2016 - 5th International Conference Youth in Conservation of Cultural heritage", Madrid, Spain, September 2016, https://hal.inria.fr/hal-01377101.

[24] J. DUMOULIN, M. MARCHETTI, N. LE TOUZ, J. M. PIAU, L. IBOS, M. FOIS.*Les matériaux à changement de phase en génie civil.Contexte spécifique de la viabilité hivernale*, in "Journée d'étude co-organisée par la Société Française de Thermique (SFT) et French Interpore Chapter sur « Réactions, échanges et stockage dans les mousses solides : Modèles et caractérisations expérimentales", Paris, France, Société Française de Thermique (SFT), March 2016, https://hal.inria.fr/hal-01377106.

International Conferences with Proceedings

[25] S. ALLAHDADIAN, M. DÖHLER, C. VENTURA, L. MEVEL.*On the influence of sample length and measurement noise on the stochastic subspace damage detection technique*, in "IMAC – 34th International Modal Analysis Conference", Orlando, FL, United States, January 2016 [*DOI :* 10.1007/978-3-319-29956-3_4], https://hal.inria.fr/hal-01262256.

[26] N. BERRABAH, M. FRANCHET, D. VAUTRIN, Q. ZHANG.*Estimation of Distributed and Lumped Ohmic Losses in Electrical Cables*, in "2016 IEEE International Conference on Antenna Measurements & Applications", Syracuse, United States, 2016 IEEE International Conference on Antenna Measurements & Applications, October 2016, https://hal.inria.fr/hal-01376833.

[27] M. D. H. BHUYAN, M. DÖHLER, L. MEVEL.*Statistical Damage Localization with Stochastic Load Vectors Using Multiple Mode Sets*, in "EWSHM - 8th European Workshop on Structural Health Monitoring", Bilbao, Spain, July 2016, https://hal.inria.fr/hal-01344206.

[28] A. CRINIÈRE, J. DUMOULIN, L. MEVEL.*Heat transfer 1D modeling for Non Destructive Testing and Structural Health Monitoring for civil engineering structures improved by GPGPU*, in "8th European Workshop On Structural Health Monitoring (EWSHM 2016)", Bilbao , Spain, 8th European Workshop On Structural Health Monitoring (EWSHM 2016), July 2016, https://hal.inria.fr/hal-01332760.

[29] A. CRINIÈRE, M. ZGHAL, L. MEVEL, J. DUMOULIN.*GPGPU implementation of modal parameter tracking by particle based Kalman filter*, in "8th European Workshop On Structural Health Monitoring (EWSHM 2016)", Bilbao , Spain, July 2016, https://hal.inria.fr/hal-01332767.

[30] M. DÖHLER, S. THÖNS.*Efficient Structural System Reliability Updating with Subspace-Based Damage Detection Information*, in "EWSHM - 8th European Workshop on Structural Health Monitoring", Bilbao, Spain, July 2016, https://hal.inria.fr/hal-01344222.

[31] G. GAUTIER, L. MEVEL, J.-M. MENCIK, M. DÖHLER, R. SERRA.*Parameter uncertainties quantification for finite element based subspace fitting approaches*, in "EWSHM - 8th European Workshop on Structural Health Monitoring", Bilbao, Spain, July 2016, https://hal.inria.fr/hal-01344198.

[32] L. IBOS, J.-P. MONCHAU, V. FEUILLET, J. DUMOULIN, P. AUSSET, J. HAMEURY, B. HAY.*Investigation of the directional emissivity of materials using infrared thermography coupled with a periodic excitation*, in "QIRT 2016", Gdańsk , Poland, July 2016 [*DOI :* 10.21611/QIRT.2016.031], https://hal.inria.fr/hal-01388188.

[33] V. LE CAM, M. DÖHLER, M. LE PEN, I. GUÉGUEN, L. MEVEL.*Embedded subspace-based modal analysis and uncertainty quantification on wireless sensor platform PEGASE*, in "EWSHM - 8th European Workshop on Structural Health Monitoring", Bilbao, Spain, July 2016, https://hal.inria.fr/hal-01344213.

[34] N. LE TOUZ, J. BROUNS, J. DUMOULIN, F. SOLDOVIERI.*Etude numérique de l'apport du couplage de deux méthodes d'identification, thermique et électromagnétique, pour la reconstruction de défauts dans des parois épaisses*, in "SFT 2016", Toulouse, France, June 2016, https://hal.inria.fr/hal-01388155.

[35] M. P. LIMONGELLI, E. CHATZI, M. DÖHLER, G. LOMBAERT, E. REYNDERS.*Towards extraction of vibration-based damage indicators*, in "EWSHM - 8th European Workshop on Structural Health Monitoring", Bilbao, Spain, July 2016, https://hal.inria.fr/hal-01344178.

[36] M. MARCHETTI, J. DUMOULIN, M. FOIS, L. IBOS, N. LE TOUZ, J.-M. PIAU.*Thermographic monitoring of asphalt concrete surface with phase change materials inclusions for icing delays purposes*, in "QIRT 2016", Gdańsk , Poland, July 2016 [*DOI :* 10.21611/QIRT.2016.078], https://hal.inria.fr/hal-01388166.

[37] P. MELLINGER, M. DÖHLER, L. MEVEL.*Variance estimation of modal parameters from input/output covariance-driven subspace identification*, in "ISMA - 27th Conference on Noise and Vibration Engineering", Leuven, Belgium, September 2016, https://hal.inria.fr/hal-01367695.

Conferences without Proceedings

[38] J. BROUNS, J. DUMOULIN.*Identification of thermal properties distribution in building wall using infrared thermography*, in "European Geosciences Union General Assembly 2016 (EGU2016)", Vienna, Austria, August 2016, vol. 18, 1 [*DOI :* 10.1117/12.2185126], https://hal.inria.fr/hal-01388147.

[39] X. CHAPELEAU, L.-M. COTTINEAU, T. SEDRAN.*"Instrumentation d'une voie ferroviaire sur dalles béton"*, in "C2I", Saint Nazaire, France, January 2016, https://hal.inria.fr/hal-01388136.

[40] A. CRINIÈRE, J. DUMOULIN, L. MEVEL, G. ANDRADE BARROSO.*Cloud2IR an infrared and environmental SHM information system*, in "13th Quantitative Infrared Thermography Conference QIRT2016", Gdansk, Poland, 13th Quantitative Infrared Thermography Conference QIRT2016, July 2016, https://hal.inria.fr/hal-01332782.

[41] A. CRINIÈRE, J. DUMOULIN, L. MEVEL, G. ANDRADE BARROSO.*Cloud2IR: Infrared thermography and environmental sensors integrated in an autonomoussystem for long term monitoring of structures*, in "European Geoscience Union EGU2016", Vienna, Austria, April 2016 [*DOI :* 10.1088/1742-2132/10/6/064003], https://hal.inria.fr/hal-01332799.

[42] A. CRINIÈRE, M. DÖHLER, V. LE CAM, L. MEVEL.*Robust uncertainty evaluation for system identification on distributed wireless platforms*, in "European Geoscience Uninon EGU 2016", Vienna, Austria, European Geoscience Uninon EGU 2016, April 2016, https://hal.inria.fr/hal-01332801.

[43] N. LE TOUZ, J. DUMOULIN, F. SOLDOVIERI.*Numerical study of the coupling of two identification methods, thermal and electromagnetic, for the reconstruction of inclusions in thick walls*, in "European Geosciences Union General Assembly", Vienne, Austria, EGU.eu, April 2016, https://hal.inria.fr/hal-01377002.

Scientific Books (or Scientific Book chapters)

[44] J. DUMOULIN.*Infrared thermography: from sensing principle to non destructive testing considerations*, in "Sensing the past", Geoscience and Sensing Technologies for cultural heritage, Springer, 2016, https://hal.inria.fr/hal-01376217.

[45] F. SOLDOVIERI, J. DUMOULIN.*Integrated monitoring at a modern architectural masterpiece: the case of Viaduct Basento in Potenza*, in "Sensing the past", Geoscience and Sensing Technologies for cultural heritage, 2016, https://hal.inria.fr/hal-01376212.

Patents and standards

[46] N. BERRABAH, Q. ZHANG.*Procédé et système de localisation de défauts sur un câble électrique*, July 2016, n^o 1657416, Enregistré par l'INPI le 29 juillet 2016, https://hal.inria.fr/hal-01405142.

References in notes

[47] M. BASSEVILLE, I. V. NIKIFOROV.*Fault isolation for diagnosis : nuisance rejection and multiple hypotheses testing*, in "Annual Reviews in Control", December 2002, vol. 26, n^o 2, p. 189–202, http://dx.doi.org/10.1016/S1367-5788(02)00029-9.

[48] B. DELYON, A. JUDITSKY, A. BENVENISTE.*On the relationship between identification and local tests*, IRISA, May 1997, n^o 1104, ftp://ftp.irisa.fr/techreports/1997/PI-1104.ps.gz.

[49] M. JAULENT.*The inverse scattering problem for LCRG transmission lines*, in "Journal of Mathematical Physics", December 1982, vol. 23, n^o 12, p. 2286-2290.

[50] G. L. LAMB.*Elements of Soliton Theory*, John Wiley & Sons, New York, 1980.

[51] M. OUMRI.*Fault diagnosis of wired electric networks by reflectometry*, Université Paris Sud - Paris XI, May 2014, https://tel.archives-ouvertes.fr/tel-01165039.

[52] C. R. PAUL.*Analysis of multiconductor transmission lines*, Wiley, New York, 2008.

[53] H. TANG, Q. ZHANG.*An Inverse Scattering Approach to Soft Fault Diagnosis in Lossy Electric Transmission Lines*, in "IEEE Trans. on Antennas and Propagation", 2011, vol. 59, n^o 10, p. 3730-3737, http://dx.doi.org/10.1109/TAP.2011.2163772.

[54] P. VAN OVERSCHEE, B. DE MOOR.*Subspace Identification for Linear Systems*, Kluwer Academic Publishers, Boston, 1996.

[55] F. VISCO COMANDINI.*Some inverse scattering problems on star-shaped graphs: application to fault detection on electrical transmission line networks*, Université de Versailles-Saint Quentin en Yvelines, December 2011, https://tel.archives-ouvertes.fr/tel-00748216.

Project-Team IPSO

Invariant Preserving SOlvers

IN COLLABORATION WITH: Institut de recherche mathématique de Rennes (IRMAR)

IN PARTNERSHIP WITH:
CNRS

Université Rennes 1

École normale supérieure de Rennes

RESEARCH CENTER
Rennes - Bretagne-Atlantique

THEME
Numerical schemes and simulations

Table of contents

Project-Team IPSO

Creation of the Project-Team: 2004 December 06

Keywords:

Computer Science and Digital Science:

6. - Modeling, simulation and control
6.1. - Mathematical Modeling
6.1.1. - Continuous Modeling (PDE, ODE)
6.1.2. - Stochastic Modeling (SPDE, SDE)
6.1.4. - Multiscale modeling
6.1.5. - Multiphysics modeling
6.2. - Scientific Computing, Numerical Analysis & Optimization
6.2.1. - Numerical analysis of PDE and ODE
6.2.2. - Numerical probability
6.2.3. - Probabilistic methods

Other Research Topics and Application Domains:

1. - Life sciences
1.1. - Biology
1.1.10. - Mathematical biology
4. - Energy
4.1. - Fossile energy production (oil, gas)
5. - Industry of the future
5.3. - Nanotechnology

1. Members

Research Scientists

Philippe Chartier [Team leader, Inria, Senior Researcher, HDR]
Nicolas Crouseilles [Inria, Researcher, HDR]
Erwan Faou [Inria, Senior Researcher, HDR]
Mohammed Lemou [CNRS, Senior Researcher, HDR]

Faculty Members

François Castella [Univ. Rennes I, Professor, HDR]
Anais Crestetto [Univ. Nantes, Associate Professor]
Arnaud Debussche [ENS Rennes, Professor, HDR]
Florian Méhats [Univ. Rennes I, Professor, HDR]

PhD Students

Hélène Hivert [Univ. Rennes I, until Aug 2016]
Romain Horsin [Inria]
Julie Sauzeau [Univ. Rennes I, until Aug 2016]
Maxime Tusseau [Univ. Rennes I, until Aug 2016]

Post-Doctoral Fellows

Nathalie Ayi [Inria]
Xiaofei Zhao [Inria, from Oct 2016]

Administrative Assistants

Stephanie Lemaile [Inria]

Armelle Mozziconacci [CNRS, from May 2016]

2. Overall Objectives

2.1. An overview of geometric numerical integration

A fundamental and enduring challenge in science and technology is the quantitative prediction of time-dependent nonlinear phenomena. While dynamical simulation (for ballistic trajectories) was one of the first applications of the digital computer, the problems treated, the methods used, and their implementation have all changed a great deal over the years. Astronomers use simulation to study long term evolution of the solar system. Molecular simulations are essential for the design of new materials and for drug discovery. Simulation can replace or guide experiment, which often is difficult or even impossible to carry out as our ability to fabricate the necessary devices is limited.

During the last decades, we have seen dramatic increases in computing power, bringing to the fore an ever widening spectrum of applications for dynamical simulation. At the boundaries of different modeling regimes, it is found that computations based on the fundamental laws of physics are under-resolved in the textbook sense of numerical methods. Because of the vast range of scales involved in modeling even relatively simple biological or material functions, this limitation will not be overcome by simply requiring more computing power within any realistic time. One therefore has to develop numerical methods which capture crucial structures even if the method is far from “converging" in the mathematical sense. In this context, we are forced increasingly to think of the numerical algorithm as a part of the modeling process itself. A major step forward in this area has been the development of structure-preserving or “geometric" integrators which maintain conservation laws, dissipation rates, or other key features of the continuous dynamical model. Conservation of energy and momentum are fundamental for many physical models; more complicated invariants are maintained in applications such as molecular dynamics and play a key role in determining the long term stability of methods. In mechanical models (biodynamics, vehicle simulation, astrodynamics) the available structure may include constraint dynamics, actuator or thruster geometry, dissipation rates and properties determined by nonlinear forms of damping.

In recent years the growth of geometric integration has been very noticeable. Features such as *symplecticity* or *time-reversibility* are now widely recognized as essential properties to preserve, owing to their physical significance. This has motivated a lot of research [53], [50], [49] and led to many significant theoretical achievements (symplectic and symmetric methods, volume-preserving integrators, Lie-group methods, ...). In practice, a few simple schemes such as the Verlet method or the Störmer method have been used for years with great success in molecular dynamics or astronomy. However, they now need to be further improved in order to fit the tremendous increase of complexity and size of the models.

2.2. Overall objectives

To become more specific, the project *IPSO* aims at finding and implementing new structure-preserving schemes and at understanding the behavior of existing ones for the following type of problems:

- systems of differential equations posed on a manifold.
- systems of differential-algebraic equations of index 2 or 3, where the constraints are part of the equations.
- Hamiltonian systems and constrained Hamiltonian systems (which are special cases of the first two items though with some additional structure).
- highly-oscillatory systems (with a special focus of those resulting from the Schrödinger equation).

Although the field of application of the ideas contained in geometric integration is extremely wide (e.g. robotics, astronomy, simulation of vehicle dynamics, biomechanical modeling, biomolecular dynamics, geodynamics, chemistry...), *IPSO* will mainly concentrate on applications for *molecular dynamics simulation* and *laser simulation*:

- There is a large demand in biomolecular modeling for models that integrate microscopic molecular dynamics simulation into statistical macroscopic quantities. These simulations involve huge systems of ordinary differential equations over very long time intervals. This is a typical situation where the determination of accurate trajectories is out of reach and where one has to rely on the good qualitative behavior of structure-preserving integrators. Due to the complexity of the problem, more efficient numerical schemes need to be developed.
- The demand for new models and/or new structure-preserving schemes is also quite large in laser simulations. The propagation of lasers induces, in most practical cases, several well-separated scales: the intrinsically highly-oscillatory *waves* travel over long distances. In this situation, filtering the oscillations in order to capture the long-term trend is what is required by physicists and engineers.

3. Research Program

3.1. Structure-preserving numerical schemes for solving ordinary differential equations

Participants: François Castella, Philippe Chartier, Erwan Faou.

ordinary differential equation, numerical integrator, invariant, Hamiltonian system, reversible system, Lie-group system

In many physical situations, the time-evolution of certain quantities may be written as a Cauchy problem for a differential equation of the form

$$\begin{aligned} y'(t) &= f(y(t)), \\ y(0) &= y_0. \end{aligned} \tag{42}$$

For a given y_0, the solution $y(t)$ at time t is denoted $\varphi_t(y_0)$. For fixed t, φ_t becomes a function of y_0 called the *flow* of (1). From this point of view, a numerical scheme with step size h for solving (1) may be regarded as an approximation Φ_h of φ_h. One of the main questions of *geometric integration* is whether *intrinsic* properties of φ_t may be passed on to Φ_h.

This question can be more specifically addressed in the following situations:

3.1.1. Reversible ODEs

The system (1) is said to be ρ-reversible if there exists an involutive linear map ρ such that

$$\rho \circ \varphi_t = \varphi_t^{-1} \circ \rho = \varphi_{-t} \circ \rho. \tag{43}$$

It is then natural to require that Φ_h satisfies the same relation. If this is so, Φ_h is said to be *symmetric*. Symmetric methods for reversible systems of ODEs are just as much important as *symplectic* methods for Hamiltonian systems and offer an interesting alternative to symplectic methods.

3.1.2. ODEs with an invariant manifold

The system (1) is said to have an invariant manifold g whenever

$$\mathcal{M} = \{y \in \mathbb{R}^n; g(y) = 0\} \tag{44}$$

is kept *globally* invariant by φ_t. In terms of derivatives and for sufficiently differentiable functions f and g, this means that

$$\forall\, y \in \mathcal{M},\; g'(y)f(y) = 0.$$

As an example, we mention Lie-group equations, for which the manifold has an additional group structure. This could possibly be exploited for the space-discretisation. Numerical methods amenable to this sort of problems have been reviewed in a recent paper [48] and divided into two classes, according to whether they use g explicitly or through a projection step. In both cases, the numerical solution is forced to live on the manifold at the expense of some Newton's iterations.

3.1.3. Hamiltonian systems

Hamiltonian problems are ordinary differential equations of the form:

$$\begin{array}{rclcl} \dot{p}(t) & = & -\nabla_q H(p(t), q(t)) & \in & \mathbb{R}^d \\ \dot{q}(t) & = & \nabla_p H(p(t), q(t)) & \in & \mathbb{R}^d \end{array} \tag{45}$$

with some prescribed initial values $(p(0), q(0)) = (p_0, q_0)$ and for some scalar function H, called the Hamiltonian. In this situation, H is an invariant of the problem. The evolution equation (4) can thus be regarded as a differential equation on the manifold

$$\mathcal{M} = \{(p, q) \in \mathbb{R}^d \times \mathbb{R}^d; H(p, q) = H(p_0, q_0)\}.$$

Besides the Hamiltonian function, there might exist other invariants for such systems: when there exist d invariants in involution, the system (4) is said to be *integrable*. Consider now the parallelogram P originating from the point $(p, q) \in \mathbb{R}^{2d}$ and spanned by the two vectors $\xi \in \mathbb{R}^{2d}$ and $\eta \in \mathbb{R}^{2d}$, and let $\omega(\xi, \eta)$ be the sum of the *oriented* areas of the projections over the planes (p_i, q_i) of P,

$$\omega(\xi, \eta) = \xi^T J \eta,$$

where J is the *canonical symplectic* matrix

$$J = \begin{bmatrix} 0 & I_d \\ -I_d & 0 \end{bmatrix}.$$

A continuously differentiable map g from $\mathbb{R}^{2d}$ to itself is called symplectic if it preserves ω, i.e. if

$$\omega(g'(p, q)\xi, g'(p, q)\eta) = \omega(\xi, \eta).$$

A fundamental property of Hamiltonian systems is that their exact flow is symplectic. Integrable Hamiltonian systems behave in a very remarkable way: as a matter of fact, their invariants persist under small perturbations, as shown in the celebrated theory of Kolmogorov, Arnold and Moser. This behavior motivates the introduction of *symplectic* numerical flows that share most of the properties of the exact flow. For practical simulations of Hamiltonian systems, symplectic methods possess an important advantage: the error-growth as a function of time is indeed linear, whereas it would typically be quadratic for non-symplectic methods.

3.1.4. Differential-algebraic equations

Whenever the number of differential equations is insufficient to determine the solution of the system, it may become necessary to solve the differential part and the constraint part altogether. Systems of this sort are called differential-algebraic systems. They can be classified according to their index, yet for the purpose of this expository section, it is enough to present the so-called index-2 systems

$$\begin{array}{rcl} \dot{y}(t) & = & f(y(t), z(t)), \\ 0 & = & g(y(t)), \end{array} \tag{46}$$

where initial values $(y(0), z(0)) = (y_0, z_0)$ are given and assumed to be consistent with the constraint manifold. By constraint manifold, we imply the intersection of the manifold

$$\mathcal{M}_1 = \{y \in \mathbb{R}^n, g(y) = 0\}$$

and of the so-called hidden manifold

$$\mathcal{M}_2 = \{(y, z) \in \mathbb{R}^n \times \mathbb{R}^m, \frac{\partial g}{\partial y}(y) f(y, z) = 0\}.$$

This manifold $\mathcal{M} = \mathcal{M}_1 \bigcap \mathcal{M}_2$ is the manifold on which the exact solution $(y(t), z(t))$ of (5) lives.

There exists a whole set of schemes which provide a numerical approximation lying on $\mathcal{M}_1$. Furthermore, this solution can be projected on the manifold $\mathcal{M}$ by standard projection techniques. However, it it worth mentioning that a projection destroys the symmetry of the underlying scheme, so that the construction of a symmetric numerical scheme preserving $\mathcal{M}$ requires a more sophisticated approach.

3.2. Highly-oscillatory systems

Participants: François Castella, Philippe Chartier, Nicolas Crouseilles, Erwan Faou, Florian Méhats, Mohammed Lemou.

second-order ODEs, oscillatory solutions, Schrödinger and wave equations, step size restrictions.

In applications to molecular dynamics or quantum dynamics for instance, the right-hand side of (1) involves *fast* forces (short-range interactions) and *slow* forces (long-range interactions). Since *fast* forces are much cheaper to evaluate than *slow* forces, it seems highly desirable to design numerical methods for which the number of evaluations of slow forces is not (at least not too much) affected by the presence of fast forces.

A typical model of highly-oscillatory systems is the second-order differential equations

$$\ddot{q} \quad = \quad -\nabla V(q) \tag{47}$$

where the potential $V(q)$ is a sum of potentials $V = W + U$ acting on different time-scales, with $\nabla^2 W$ positive definite and $\|\nabla^2 W\| >> \|\nabla^2 U\|$. In order to get a bounded error propagation in the linearized equations for an explicit numerical method, the step size must be restricted according to

$$h\omega < C,$$

where C is a constant depending on the numerical method and where ω is the highest frequency of the problem, i.e. in this situation the square root of the largest eigenvalue of $\nabla^2 W$. In applications to molecular dynamics for instance, *fast* forces deriving from W (short-range interactions) are much cheaper to evaluate than *slow* forces deriving from U (long-range interactions). In this case, it thus seems highly desirable to design numerical methods for which the number of evaluations of slow forces is not (at least not too much) affected by the presence of fast forces.

Another prominent example of highly-oscillatory systems is encountered in quantum dynamics where the Schrödinger equation is the model to be used. Assuming that the Laplacian has been discretized in space, one indeed gets the *time*-dependent Schrödinger equation:

$$i\dot{\psi}(t) = \frac{1}{\varepsilon} H(t)\psi(t), \tag{48}$$

where $H(t)$ is finite-dimensional matrix and where ε typically is the square-root of a mass-ratio (say electron/ion for instance) and is small ($\varepsilon \approx 10^{-2}$ or smaller). Through the coupling with classical mechanics ($H(t)$ is obtained by solving some equations from classical mechanics), we are faced once again with two different time-scales, 1 and ε. In this situation also, it is thus desirable to devise a numerical method able to advance the solution by a time-step $h > \varepsilon$.

3.3. Geometric schemes for the Schrödinger equation

Participants: François Castella, Philippe Chartier, Erwan Faou, Florian Méhats.

Schrödinger equation, variational splitting, energy conservation.

Given the Hamiltonian structure of the Schrödinger equation, we are led to consider the question of energy preservation for time-discretization schemes.

At a higher level, the Schrödinger equation is a partial differential equation which may exhibit Hamiltonian structures. This is the case of the time-dependent Schrödinger equation, which we may write as

$$i\varepsilon \frac{\partial \psi}{\partial t} = H\psi, \tag{49}$$

where $\psi = \psi(x,t)$ is the wave function depending on the spatial variables $x = (x_1, \cdots, x_N)$ with $x_k \in \mathbb{R}^d$ (e.g., with $d = 1$ or 3 in the partition) and the time $t \in \mathbb{R}$. Here, ε is a (small) positive number representing the scaled Planck constant and i is the complex imaginary unit. The Hamiltonian operator H is written

$$H = T + V$$

with the kinetic and potential energy operators

$$T = -\sum_{k=1}^{N} \frac{\varepsilon^2}{2m_k} \Delta_{x_k} \quad \text{and} \quad V = V(x),$$

where $m_k > 0$ is a particle mass and Δ_{x_k} the Laplacian in the variable $x_k \in \mathbb{R}^d$, and where the real-valued potential V acts as a multiplication operator on ψ.

The multiplication by i in (8) plays the role of the multiplication by J in classical mechanics, and the energy $\langle \psi | H | \psi \rangle$ is conserved along the solution of (8), using the physicists' notations $\langle u | A | u \rangle = \langle u, Au \rangle$ where $\langle \ , \ \rangle$ denotes the Hermitian L^2-product over the phase space. In quantum mechanics, the number N of particles is very large making the direct approximation of (8) very difficult.

The numerical approximation of (8) can be obtained using projections onto submanifolds of the phase space, leading to various PDEs or ODEs: see [52], [51] for reviews. However the long-time behavior of these approximated solutions is well understood only in this latter case, where the dynamics turns out to be finite dimensional. In the general case, it is very difficult to prove the preservation of qualitative properties of (8) such as energy conservation or growth in time of Sobolev norms. The reason for this is that backward error analysis is not directly applicable for PDEs. Overwhelming these difficulties is thus a very interesting challenge.

A particularly interesting case of study is given by symmetric splitting methods, such as the Strang splitting:

$$\psi_1 = \exp\left(-i(\delta t)V/2\right)\exp\left(i(\delta t)\Delta\right)\exp\left(-i(\delta t)V/2\right)\psi_0 \tag{50}$$

where δt is the time increment (we have set all the parameters to 1 in the equation). As the Laplace operator is unbounded, we cannot apply the standard methods used in ODEs to derive long-time properties of these schemes. However, its projection onto finite dimensional submanifolds (such as Gaussian wave packets space or FEM finite dimensional space of functions in x) may exhibit Hamiltonian or Poisson structure, whose long-time properties turn out to be more tractable.

3.4. High-frequency limit of the Helmholtz equation

Participant: François Castella.

waves, Helmholtz equation, high oscillations.

The Helmholtz equation models the propagation of waves in a medium with variable refraction index. It is a simplified version of the Maxwell system for electro-magnetic waves.

The high-frequency regime is characterized by the fact that the typical wavelength of the signals under consideration is much smaller than the typical distance of observation of those signals. Hence, in the high-frequency regime, the Helmholtz equation at once involves highly oscillatory phenomena that are to be described in some asymptotic way. Quantitatively, the Helmholtz equation reads

$$i\alpha_\varepsilon u_\varepsilon(x) + \varepsilon^2 \Delta_x u_\varepsilon + n^2(x) u_\varepsilon = f_\varepsilon(x). \tag{51}$$

Here, ε is the small adimensional parameter that measures the typical wavelength of the signal, $n(x)$ is the space-dependent refraction index, and $f_\varepsilon(x)$ is a given (possibly dependent on ε) source term. The unknown is $u_\varepsilon(x)$. One may think of an antenna emitting waves in the whole space (this is the $f_\varepsilon(x)$), thus creating at any point x the signal $u_\varepsilon(x)$ along the propagation. The small $\alpha_\varepsilon > 0$ term takes into account damping of the waves as they propagate.

One important scientific objective typically is to describe the high-frequency regime in terms of *rays* propagating in the medium, that are possibly refracted at interfaces, or bounce on boundaries, etc. Ultimately, one would like to replace the true numerical resolution of the Helmholtz equation by that of a simpler, asymptotic model, formulated in terms of rays.

In some sense, and in comparison with, say, the wave equation, the specificity of the Helmholtz equation is the following. While the wave equation typically describes the evolution of waves between some initial time and some given observation time, the Helmholtz equation takes into account at once the propagation of waves over *infinitely long* time intervals. Qualitatively, in order to have a good understanding of the signal observed in some bounded region of space, one readily needs to be able to describe the propagative phenomena in the whole space, up to infinity. In other words, the "rays" we refer to above need to be understood from the initial time up to infinity. This is a central difficulty in the analysis of the high-frequency behaviour of the Helmholtz equation.

3.5. From the Schrödinger equation to Boltzmann-like equations

Participant: François Castella.

Schrödinger equation, asymptotic model, Boltzmann equation.

The Schrödinger equation is the appropriate way to describe transport phenomena at the scale of electrons. However, for real devices, it is important to derive models valid at a larger scale.

In semi-conductors, the Schrödinger equation is the ultimate model that allows to obtain quantitative information about electronic transport in crystals. It reads, in convenient adimensional units,

$$i\partial_t \psi(t,x) = -\frac{1}{2}\Delta_x \psi + V(x)\psi, \tag{52}$$

where $V(x)$ is the potential and $\psi(t,x)$ is the time- and space-dependent wave function. However, the size of real devices makes it important to derive simplified models that are valid at a larger scale. Typically, one wishes to have kinetic transport equations. As is well-known, this requirement needs one to be able to describe "collisions" between electrons in these devices, a concept that makes sense at the macroscopic level, while it does not at the microscopic (electronic) level. Quantitatively, the question is the following: can one obtain the Boltzmann equation (an equation that describes collisional phenomena) as an asymptotic model for the Schrödinger equation, along the physically relevant micro-macro asymptotics? From the point of view of modelling, one wishes here to understand what are the "good objects", or, in more technical words, what are the relevant "cross-sections", that describe the elementary collisional phenomena. Quantitatively, the Boltzmann equation reads, in a simplified, linearized, form :

$$\partial_t f(t,x,v) = \int_{\mathbf{R}^3} \sigma(v,v')\,[f(t,x,v') - f(t,x,v)]dv'. \tag{53}$$

Here, the unknown is $f(x,v,t)$, the probability that a particle sits at position x, with a velocity v, at time t. Also, $\sigma(v,v')$ is called the cross-section, and it describes the probability that a particle "jumps" from velocity v to velocity v' (or the converse) after a collision process.

4. New Results

4.1. List of results

4.1.1. Landau damping in Sobolev spaces for the Vlasov-HMF model

In [25], the authors consider the Vlasov-HMF (Hamiltonian Mean-Field) model. They consider solutions starting in a small Sobolev neighborhood of a spatially homogeneous state satisfying a linearized stability criterion (Penrose criterion). They prove that these solutions exhibit a scattering behavior to a modified state, which implies a nonlinear Landau damping effect with polynomial rate of damping.

4.1.2. Fast Weak-Kam Integrators for separable Hamiltonian systems

In [4], the authors consider a numerical scheme for Hamilton-Jacobi equations based on a direct discretization of the Lax-Oleinik semi-group. They prove that this method is convergent with respect to the time and space stepsizes provided the solution is Lipschitz, and give an error estimate. Moreover, They prove that the numerical scheme is a *geometric integrator* satisfying a discrete weak-KAM theorem which allows to control its long time behavior. Taking advantage of a fast algorithm for computing min–plus convolutions based on the decomposition of the function into concave and convex parts, they show that the numerical scheme can be implemented in a very efficient way.

4.1.3. The weakly nonlinear large-box limit of the 2D cubic nonlinear Schrödinger equation

In [23], the authors consider the cubic nonlinear Schrödinger (NLS) equation set on a two dimensional box of size L with periodic boundary conditions. By taking the large box limit $L \to \infty$ in the weakly nonlinear regime (characterized by smallness in the critical space), we derive a new equation set on $\mathbb{R}^2$ that approximates the dynamics of the frequency modes. This nonlinear equation turns out to be Hamiltonian and enjoys interesting symmetries, such as its invariance under Fourier transform, as well as several families of explicit solutions. A large part of this work is devoted to a rigorous approximation result that allows to project the long-time dynamics of the limit equation into that of the cubic NLS equation on a box of finite size.

4.1.4. An asymptotic preserving scheme for the relativistic Vlasov–Maxwell equations in the classical limit

In [13], the authors consider the relativistic Vlasov–Maxwell (RVM) equations in the limit when the light velocity c goes to infinity. In this regime, the RVM system converges towards the Vlasov–Poisson system and the aim of this work is to construct asymptotic preserving numerical schemes that are robust with respect to this limit. A number of numerical simulations are conducted in order to investigate the performances of our numerical scheme both in the relativistic as well as in the classical limit regime. In addition, they derive the dispersion relation of the Weibel instability for the continuous and the discretized problem.

4.1.5. Free Vibrations of Axisymmetric Shells: Parabolic and Elliptic cases

In [41], approximate eigenpairs (quasimodes) of axisymmetric thin elastic domains with laterally clamped boundary conditions (Lamé system) are determined by an asymptotic analysis as the thickness (2ε) tends to zero. The departing point is the Koiter shell model that we reduce by asymptotic analysis to a scalar model that depends on two parameters: the angular frequency k and the half-thickness ε. Optimizing k for each chosen ε, we find power laws for k in function of ε that provide the smallest eigenvalues of the scalar reductions. Corresponding eigenpairs generate quasimodes for the 3D Lamé system by means of several reconstruction operators, including boundary layer terms. Numerical experiments demonstrate that in many cases the constructed eigenpair corresponds to the first eigenpair of the Lamé system.

Geometrical conditions are necessary to this approach: The Gaussian curvature has to be nonnegative and the azimuthal curvature has to dominate the meridian curvature in any point of the midsurface. In this case, the first eigenvector admits progressively larger oscillation in the angular variable as ε tends to 0. Its angular frequency exhibits a power law relation of the form $k = \gamma\varepsilon^{-\beta}$ with $\beta = \frac{1}{4}$ in the parabolic case (cylinders and trimmed cones), and the various βs $\frac{2}{5}$, $\frac{3}{7}$, and $\frac{1}{3}$ in the elliptic case. For these cases where the mathematical analysis is applicable, numerical examples that illustrate the theoretical results are presented.

4.1.6. High frequency oscillations of first eigenmodes in axisymmetric shells as the thickness tends to zero

In [30], the lowest eigenmode of thin axisymmetric shells is investigated for two physical models (acoustics and elasticity) as the shell thickness (2ε) tends to zero. Using a novel asymptotic expansion we determine the behavior of the eigenvalue $\lambda(\varepsilon)$ and the eigenvector angular frequency $k(\varepsilon)$ for shells with Dirichlet boundary conditions along the lateral boundary, and natural boundary conditions on the other parts.

First, the scalar Laplace operator for acoustics is addressed, for which $k(\varepsilon)$ is always zero. In contrast to it, for the Lamé system of linear elasticity several different types of shells are defined, characterized by their geometry, for which $k(\varepsilon)$ tends to infinity as ε tends to zero. For two families of shells: cylinders and elliptical barrels we explicitly provide $\lambda(\varepsilon)$ and $k(\varepsilon)$ and demonstrate by numerical examples the different behavior as ε tends to zero.

4.1.7. On numerical Landau damping for splitting methods applied to the Vlasov-HMF model

In [24], we consider time discretizations of the Vlasov-HMF (Hamiltonian Mean-Field) equation based on splitting methods between the linear and non-linear parts. We consider solutions starting in a small Sobolev neighborhood of a spatially homogeneous state satisfying a linearized stability criterion (Penrose criterion). We prove that the numerical solutions exhibit a scattering behavior to a modified state, which implies a nonlinear Landau damping effect with polynomial rate of damping. Moreover, we prove that the modified state is close to the continuous one and provide error estimates with respect to the time stepsize.

4.1.8. High-order Hamiltonian splitting for Vlasov-Poisson equations

In [5], we consider the Vlasov-Poisson equation in a Hamiltonian framework and derive new time splitting methods based on the decomposition of the Hamiltonian functional between the kinetic and electric energy. Assuming smoothness of the solutions, we study the order conditions of such methods. It appears that these conditions are of Runge-Kutta-Nyström type. In the one dimensional case, the order conditions can be further simplified, and efficient methods of order 6 with a reduced number of stages can be constructed. In the general case, high-order methods can also be constructed using explicit computations of commutators. Numerical results are performed and show the benefit of using high-order splitting schemes in that context. Complete and self-contained proofs of convergence results and rigorous error estimates are also given.

4.1.9. Uniformly accurate exponential-type integrators for Klein-Gordon equations with asymptotic convergence to classical splitting schemes in the nonlinear schrödinger limit

In [34], we introduce efficient and robust exponential-type integrators for Klein-Gordon equations which resolve the solution in the relativistic regime as well as in the highly-oscillatory non-relativistic regime without any step-size restriction under the same regularity assumptions on the initial data required for the integration of the corresponding nonlinear Schrödinger limit system. In contrast to previous works we do not employ any asymptotic or multiscale expansion of the solution. This allows us to derive uniform convergent schemes under far weaker regularity assumptions on the exact solution. In addition, the newly derived first- and second-order exponential-type integrators converge to the classical Lie, respectively, Strang splitting in the nonlinear Schrödinger limit.

4.1.10. Convergence of a normalized gradient algorithm for computing ground states

In [45], we consider the approximation of the ground state of the one-dimensional cubic nonlinear Schrödinger equation by a normalized gradient algorithm combined with linearly implicit time integrator, and finite difference space approximation. We show that this method, also called *imaginary time evolution method* in the physics literature, is locally convergent, and we provide error estimates: for an initial data in a neighborhood of the ground state, the algorithm converges exponentially towards a modified soliton that is a space discretization of the exact soliton, with error estimates depending on the discretization parameters.

4.1.11. Improved error estimates for splitting methods applied to highly-oscillatory nonlinear Schrödinger equations

In [8], we analyse the error behavior of operator splitting methods for highly-oscillatory differential equations. The scope of applications includes time-dependent nonlinear Schrödinger equations, where the evolution operator associated with the principal linear part is highly-oscillatory and periodic in time. In a first step, a known convergence result for the second-order Strang splitting method applied to the cubic Schrödinger equation is adapted to a wider class of nonlinearities. In a second step, the dependence of the global error on the decisive parameter $0 < \varepsilon < 1$, defining the length of the period, is examined. The main result states that, compared to established error estimates, the Strang splitting method is more accurate by a factor ε, provided that the time stepsize is chosen as an integer fraction of the period. This improved error behavior over a time interval of fixed length, which is independent of the period, is due to an averaging effect. The extension of the convergence result to higher-order splitting methods and numerical illustrations complement the investigations.

4.1.12. Solving highly-oscillatory NLS with SAM: numerical efficiency and geometric properties

In [7], we present the Stroboscopic Averaging Method (SAM), which aims at numerically solving highly-oscillatory differential equations. More specifically, we first apply SAM to the Schrödinger equation on the 1-dimensional torus and on the real line with harmonic potential, with the aim of assessing its efficiency: as compared to the well-established standard splitting schemes, the stiffer the problem is, the larger the speed-up grows (up to a factor 100 in our tests). The geometric properties of SAM are also explored: on very long time intervals, symmetric implementations of the method show a very good preservation of the mass invariant and of the energy. In a second series of experiments on 2-dimensional equations, we demonstrate the ability of SAM to capture qualitatively the long-time evolution of the solution (without spurring high oscillations).

4.1.13. Highly-oscillatory evolution equations with non-resonant frequencies: averaging and numerics

In [40], we are concerned with the application of the recently introduced multi-revolution composition methods, on the one hand, and two-scale methods, on the other hand, to a class of highly-oscillatory evolution equations with multiple frequencies. The main idea relies on a well-balanced reformulation of the problem as an equivalent mono-frequency equation which allows for the use of the two aforementioned techniques.

4.1.14. A formal series approach to the Center Manifold theorem

In [35], we consider near-equilibrium systems of ordinary differential equations with explicit separation of the slow and stable manifolds. Formal B-series like those previously used to analyze highly-oscillatory systems or to construct modified equations are employed here to construct expansions of the change of variables, the center invariant manifold and the reduced model. The new approach may be seen as a process of reduction to a normal form, with the main advantage, as compared to the standard view conveyed by the celebrated center manifold theorem, that it is possible to recover the complete solution at any time through an explicit change of variables.

4.1.15. Uniformly accurate time-splitting methods for the semi-classical Schrödinger equation, Part II: Numerical analysis

This article [39] is second part of a twofold paper devoted to the construction of numerical methods which remain insensitive to the smallness of the semiclassical parameter for the Schrödinger equation in the semiclassical limit. Here, we specifically analyse the convergence behavior of the first-order splitting introduced in Part I, for a linear equation with smooth potential. Our main result is a proof of uniform accuracy.

4.1.16. Averaging of highly-oscillatory transport equations

In [38], we develop a new strategy aimed at obtaining high-order asymptotic models for transport equations with highly-oscillatory solutions. The technique relies upon recent developments averaging theory for ordinary differential equations, in particular normal form expansions in the vanishing parameter. Noteworthy, the result we state here also allows for the complete recovery of the exact solution from the asymptotic model. This is done by solving a companion transport equation that stems naturally from the change of variables underlying high-order averaging. Eventually, we apply our technique to the Vlasov equation with external electric and magnetic fields. Both constant and non-constant magnetic fields are envisaged, and asymptotic models already documented in the literature and re-derived using our methodology. In addition, it is shown how to obtain new high-order asymptotic models.

4.1.17. Asymptotic preserving and time diminishing schemes for rarefied gas dynamic

In [11], we introduce a new class of numerical schemes for rarefied gas dynamic problems described by collisional kinetic equations. The idea consists in reformulating the problem using a micro-macro decomposition and successively in solving the microscopic part by using asymptotically stable Monte Carlo methods. We consider two types of decompositions, the first leading to the Euler system of gas dynamics while the second to the Navier-Stokes equations for the macroscopic part. In addition, the particle method which solves the microscopic part is designed in such a way that the global scheme becomes computationally less expensive as the solution approaches the equilibrium state as opposite to standard methods for kinetic equations which computational cost increases with the number of interactions. At the same time, the statistical error due to the particle part of the solution decreases as the system approach the equilibrium state. This causes the method to degenerate to the sole solution of the macroscopic hydrodynamic equations (Euler or Navier-Stokes) in the limit of infinite number of collisions. In a last part, we will show the behaviors of this new approach in comparisons to standard Monte Carlo techniques for solving the kinetic equation by testing it on different problems which typically arise in rarefied gas dynamic simulations.

4.1.18. Asymptotic Preserving scheme for a kinetic model describing incompressible fluids

The kinetic theory of fluid turbulence modeling developed by Degond and Lemou (2002) 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 in this work [17] 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.

4.1.19. Numerical schemes for kinetic equations in the diffusion and anomalous diffusion limits. Part I: the case of heavy-tailed equilibrium

In [15], we propose some numerical schemes for linear kinetic equations in the diffusion and anomalous diffusion limit. When the equilibrium distribution function is a Maxwellian distribution, it is well known that for an appropriate time scale, the small mean free path limit gives rise to a diffusion type equation. However, when a heavy-tailed distribution is considered, another time scale is required and the small mean free path limit leads to a fractional anomalous diffusion equation. Our aim is to develop numerical schemes for the original kinetic model which works for the different regimes, without being restricted by stability conditions of standard explicit time integrators. First, we propose some numerical schemes for the diffusion asymptotics; then, their extension to the anomalous diffusion limit is studied. In this case, it is crucial to capture the effect of the large velocities of the heavy-tailed equilibrium, so that some important transformations of the schemes derived for the diffusion asymptotics are needed. As a result, we obtain numerical schemes which enjoy the Asymptotic Preserving property in the anomalous diffusion limit, that is: they do not suffer from the restriction on the time step and they degenerate towards the fractional diffusion limit when the mean free path goes to zero. We also numerically investigate the uniform accuracy and construct a class of numerical schemes satisfying this property. Finally, the efficiency of the different numerical schemes is shown through numerical experiments.

4.1.20. Numerical schemes for kinetic equations in the anomalous diffusion limit. Part II: degenerate collision frequency

In [14], which is the continuation of [15], we propose numerical schemes for linear kinetic equation which are able to deal with the fractional diffusion limit. When the collision frequency degenerates for small velocities it is known that for an appropriate time scale, the small mean free path limit leads to an anomalous diffusion equation. From a numerical point of view, this degeneracy gives rise to an additional stiffness that must be treated in a suitable way to avoid a prohibitive computational cost. Our aim is therefore to construct a class of numerical schemes which are able to undertake these stiffness. This means that the numerical schemes are able to capture the effect of small velocities in the small mean free path limit with a fixed set of numerical parameters. Various numerical tests are performed to illustrate the efficiency of our methods in this context.

4.1.21. Multiscale schemes for the BGK-Vlasov-Poisson system in the quasi-neutral and fluid limits. Stability analysis and first order schemes

In [12], 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.

4.1.22. Uniformly accurate forward semi-Lagrangian methods for highly oscillatory Vlasov-Poisson equations.

In [16], we deal with the numerical simulation of a Vlasov-Poisson equation modeling charged particles in a beam submitted to a highly oscillatory external electric field. A numerical scheme is constructed for this model. This scheme is uniformly accurate with respect to the size of the fast time oscillations of the solution, which means that no time step refinement is required to simulate the problem. The scheme combines the forward semi-Lagrangian method with a class of Uniformly Accurate (UA) time integrators to solve the characteristics. These UA time integrators are derived by means of a two-scale formulation of the characteristics, with the introduction of an additional periodic variable. Numerical experiments are done to show the efficiency of the proposed methods compared to conventional approaches.

4.1.23. Multi-scale methods for the solution of the radiative transfer equation

Various methods have been developed and tested over the years to solve the radiative transfer equation (RTE) with different results and trade-offs. Although the RTE is extensively used, the approximate diffusion equation is sometimes preferred, particularly in optically thick media, due to the lower computational requirements. Recently, multi-scale models, namely the domain decomposition methods, the micro-macro model and the hybrid transport- diffusion model, have been proposed as an alternative to the RTE. In domain decomposition methods, the domain is split into two subdomains, namely a mesoscopic subdomain where the RTE is solved and a macroscopic subdomain where the diffusion equation is solved. In the micro-macro and hybrid transport-diffusion models, the radiation intensity is decomposed into a macroscopic component and a mesoscopic one. In both cases, the aim is to reduce the computational requirements, while maintaining the accuracy, or to improve the accuracy for similar computational requirements. In [10], these multi-scale methods are described, and the application of the micro-macro and hybrid transport-diffusion models to a three- dimensional transient problem is reported. It is shown that when the diffusion approximation is accurate, but not over the entire domain, the multi-scale methods may improve the solution accuracy in comparison with the solution of the RTE. The order of accuracy of the numerical schemes and the radiative properties of the medium play a key role in the performance of the multi-scale methods.

4.1.24. Nonlinear Geometric Optics method based multi-scale numerical schemes for highly-oscillatory transport equations

In [42], we introduce a new numerical strategy to solve a class of oscillatory transport PDE models which is able to capture accurately the solutions without numerically resolving the high frequency oscillations in both space and time. Such PDE models arise in semiclassical modeling of quantum dynamics with band-crossings, and other highly oscillatory waves. Our first main idea is to use the nonlinear geometric optics ansatz, which builds the oscillatory phase into an independent variable. We then choose suitable initial data, based on the Chapman-Enskog expansion, for the new model. For a scalar model, we prove that so constructed model will have certain smoothness, and consequently, for a first order approximation scheme we prove uniform error estimates independent of the (possibly small) wave length. The method is extended to systems arising from a semiclassical model for surface hopping, a non-adiabatic quantum dynamic phenomenon. Numerous numerical examples demonstrate that the method has the desired properties.

4.1.25. Asymptotic Preserving numerical schemes for multiscale parabolic problems

In [18], we consider a class of multiscale parabolic problems with diffusion coefficients oscillating in space at a possibly small scale ε. Numerical homogenization methods are popular for such problems, because they capture efficiently the asymptotic behaviour as ε goes to 0, without using a dramatically fine spatial discretization at the scale of the fast oscillations. However, known such homogenization schemes are in general not accurate for both the highly oscillatory regime ($\varepsilon << 1$) and the non oscillatory regime ($\varepsilon \approx 1$). In this paper, we introduce an Asymptotic Preserving method based on an exact micro-macro decomposition of the solution which remains consistent for both regimes.

4.1.26. Uniformly accurate numerical schemes for the nonlinear dirac equation in the nonrelativistic limit regime

In [47], we apply the two-scale formulation approach to propose uniformly accurate (UA) schemes for solving the nonlinear Dirac equation in the nonrelativistic limit regime. The nonlinear Dirac equation involves two small scales ε and ε^2 with $\varepsilon \to 0$ in the nonrelativistic limit regime. The small parameter causes high oscillations in time which bring severe numerical burden for classical numerical methods. We present a suitable two-scale formulation as a general strategy to tackle a class of highly oscillatory problems involving the two small scales ε and ε^2. A numerical scheme with uniform (with respect to $\varepsilon \in [0,1]$) second order accuracy in time and a spectral accuracy in space are proposed. Numerical experiments are done to confirm the UA property.

4.1.27. Semiclassical Sobolev constants for the electro-magnetic Robin Laplacian

In [26], we deal with the asymptotic analysis of the optimal Sobolev constants in the semiclassical limit and in any dimension. We combine semiclassical arguments and concentration-compactness estimates to tackle the case when an electromagnetic field is added as well as a smooth boundary carrying a Robin condition. As a byproduct of the semiclassical strategy, we also get exponentially weighted localization estimates of the minimizers.

4.1.28. On the MIT bag model: self-adjointness and non-relativistic limit

This paper [32] is devoted to the mathematical investigation of the MIT bag model, that is the Dirac operator on a smooth and bounded domain with certain boundary conditions. We prove that the operator is self-adjoint and, when the mass goes to infinity, we provide spectral asymptotic results.

4.1.29. Global behavior of N competing species with strong diffusion: diffusion leads to exclusion

It is known that the competitive exclusion principle holds for a large kind of models involving several species competing for a single resource in an homogeneous environment. Various works indicate that the coexistence is possible in an heterogeneous environment. We propose in [6] a spatially heterogeneous system modeling the competition of several species for a single resource. If spatial movements are fast enough, we show that our system can be well approximated by a spatially homogeneous system, called aggregated model, which can be explicitly computed. Moreover, we show that if the competitive exclusion principle holds for the aggregated model, it holds for the spatially heterogeneous model too.

4.1.30. Extended Rearrangement inequalities and applications to some quantitative stability results

In [28], we prove a new functional inequality of Hardy-Littlewood type for generalized rearrangements of functions. We then show how this inequality provides *quantitative* stability results of steady states to evolution systems that essentially preserve the rearrangements and some suitable energy functional, under minimal regularity assumptions on the perturbations. In particular, this inequality yields a *quantitative* stability result of a large class of steady state solutions to the Vlasov-Poisson systems, and more precisely we derive a quantitative control of the L^1 norm of the perturbation by the relative Hamiltonian (the energy functional) and rearrangements. A general non linear stability result has been obtained recently by Lemou, Méhats and Raphaël (2012) in the gravitational context, however the proof relied in a crucial way on compactness arguments which by construction provides no quantitative control of the perturbation. Our functional inequality is also applied to the context of 2D-Euler system and also provides quantitative stability results of a large class of steady-states to this system in a natural energy space.

4.1.31. Mate Finding, Sexual Spore Production, and the Spread of Fungal Plant Parasites

Sexual reproduction and dispersal are often coupled in organisms mixing sexual and asexual reproduction, such as fungi. The aim of this study [27] is to evaluate the impact of mate limitation on the spreading speed of fungal plant parasites. Starting from a simple model with two coupled partial differential equations, we take advantage of the fact that we are interested in the dynamics over large spatial and temporal scales to reduce the model to a single equation. We obtain a simple expression for speed of spread, accounting for both sexual and asexual reproduction. Taking Black Sigatoka disease of banana plants as a case study, the model prediction is in close agreement with the actual spreading speed (100 km per year), whereas a similar model without mate limitation predicts a wave speed one order of magnitude greater. We discuss the implications of these results to control parasites in which sexual reproduction and dispersal are intrinsically coupled.

4.1.32. Dimension reduction for rotating Bose-Einstein condensates with anisotropic confinement

In [29], we consider the three-dimensional time-dependent Gross-Pitaevskii equation arising in the description of rotating Bose-Einstein condensates and study the corresponding scaling limit of strongly anisotropic confinement potentials. The resulting effective equations in one or two spatial dimensions, respectively, are rigorously obtained as special cases of an averaged three dimensional limit model. In the particular case where the rotation axis is not parallel to the strongly confining direction the resulting limiting model(s) include a negative, and thus, purely repulsive quadratic potential, which is not present in the original equation and which can be seen as an effective centrifugal force counteracting the confinement.

4.1.33. Averaging of nonlinear Schrödinger equations with strong magnetic confinement

In [46], we consider the dynamics of nonlinear Schrödinger equations with strong constant magnetic fields. In an asymptotic scaling limit the system exhibits a purely magnetic confinement, based on the spectral properties of the Landau Hamiltonian. Using an averaging technique we derive an associated effective description via an averaged model of nonlinear Schrödinger type. In a special case this also yields a derivation of the LLL equation.

4.1.34. The Interaction Picture method for solving the generalized nonlinear Schrödinger equation in optics

The interaction picture (IP) method is a very promising alternative to Split-Step methods for solving certain type of partial differential equations such as the nonlinear Schrödinger equation used in the simulation of wave propagation in optical fibers. The method exhibits interesting convergence properties and is likely to provide more accurate numerical results than cost comparable Split-Step methods such as the Symmetric Split-Step method. In [1] we investigate in detail the numerical properties of the IP method and carry out a precise comparison between the IP method and the Symmetric Split-Step method.

4.1.35. Diffusion limit for the radiative transfer equation perturbed by a Markovian process

In [21], 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.

4.1.36. Estimate for P_tD for the stochastic Burgers equation

In [20], we consider the Burgers equation on $H = L^2(0,1)$ perturbed by white noise and the corresponding transition semigroup P_t. We prove a new formula for $P_tD\phi$ (where $\phi : H \to \mathbb{R}$ is bounded and Borel) which depends on ϕ but not on its derivative. Then we deduce some new consequences for the invariant measure ν of P_t as its Fomin differentiability and an integration by parts formula which generalises the classical one for gaussian measures.

4.1.37. Degenerate Parabolic Stochastic Partial Differential Equations: Quasilinear case

In [22], we study the Cauchy problem for a quasilinear degenerate parabolic stochastic partial differential equation driven by a cylindrical Wiener process. In particular, we adapt the notion of kinetic formulation and kinetic solution and develop a well-posedness theory that includes also an L^1-contraction property. In comparison to the first-order case (Debussche and Vovelle, 2010) and to the semilinear degenerate parabolic case (Hofmanová, 2013), the present result contains two new ingredients: a generalized Itô formula that permits a rigorous derivation of the kinetic formulation even in the case of weak solutions of certain nondegenerate approximations and a direct proof of strong convergence of these approximations to the desired kinetic solution of the degenerate problem.

4.1.38. An integral inequality for the invariant measure of a stochastic reaction-diffusion equation

In [19], we consider a reaction-diffusion equation perturbed by noise (not necessarily white). We prove an integral inequality for the invariant measure ν of a stochastic reaction-diffusion equation. Then we discuss some consequences as an integration by parts formula which extends to ν a basic identity of the Malliavin Calculus. Finally, we prove the existence of a surface measure for a ball and a half-space of H.

4.1.39. Large deviations for the two-dimensional stochastic Navier-Stokes equation with vanishing noise correlation

In [36], we are dealing with the validity of a large deviation principle for the two-dimensional Navier-Stokes equation, with periodic boundary conditions, perturbed by a Gaussian random forcing. We are here interested in the regime where both the strength of the noise and its correlation are vanishing, on a length scale ε and $\delta(\varepsilon)$, respectively, with $0 < \varepsilon, \delta(\varepsilon) << 1$. Depending on the relationship between ε and $\delta(\varepsilon)$ we will prove the validity of the large deviation principle in different functional spaces.

4.1.40. Quasilinear generalized parabolic Anderson model

In [33], we provide a local in time well-posedness result for a quasilinear generalized parabolic Anderson model in dimension two $\partial_t u = a(u)\Delta u + g(u)\xi$. The key idea of our approach is a simple transformation of the equation which allows to treat the problem as a semilinear problem. The analysis is done within the setting of paracontrolled calculus.

4.1.41. The Schrödinger equation with spatial white noise potential

In [44], we consider the linear and nonlinear Schrödinger equation with a spatial white noise as a potential in dimension 2. We prove existence and uniqueness of solutions thanks to a change of unknown used by Hairer and Labbé (2015) and conserved quantities.

5. Partnerships and Cooperations

5.1. National Initiatives

5.1.1. ANR MOONRISE: 2015-2019

Participants: François Castella, Philippe Chartier, Nicolas Crouseilles, Mohammed Lemou, Florian Méhats.

The project *Moonrise* submitted by Florian Méhats has been funded by the ANR for 4 years, for the period 2015-2019. This project aims at exploring modeling, mathematical and numerical issues originating from the presence of high-oscillations in nonlinear PDEs from the physics of nanotechnologies (quantum transport) and from the physics of plasmas (magnetized transport in tokamaks). The partners of the project are the IRMAR (Rennes), the IMT (Toulouse) and the CEA Cadarache. In the IPSO team, François Castella, Philippe Chartier, Nicolas Crouseilles and Mohammed Lemou are members of the project Moonrise.

Postdocs

- Loïc Le Treust has been hired as a Postdoc, under the supervision of Philippe Chartier and Florian Méhats. His contract started in september 2015 and ended in august 2016. Loïc Le Treust is now assistant professor at the university of Marseille.
- Xiaofei Zhao has been hired as a Postdoc from september 2015 to september 2016 under the supervision of Florian Méhats.

5.1.2. ANR MFG: 2016-2020

Participant: Arnaud Debussche.

Mean Field Games (MFG) theory is a new and challenging mathematical topic which analyzes the dynamics of a very large number of interacting rational agents. Introduced ten years ago, the MFG models have been used in many areas such as, e.g., economics (heterogeneous agent models, growth modeling,...), finance (formation of volatility, models of bank runs,...), social sciences (crowd models, models of segregation) and engineering (data networks, energy systems...). Their importance comes from the fact that they are the simplest (“stochastic control"-type) models taking into account interactions between rational agents (thus getting beyond optimization), yet without entering into the issues of strategic interactions. MFG theory lies at the intersection of mean field theories (it studies systems with a very large number of agents), game theory, optimal control and stochastic analysis (the agents optimize a payoff in a possibly noisy setting), calculus of variations (MFG equilibria may arise as minima of suitable functionals) and partial differential equations (PDE): In the simplest cases, the value of each agent is found by solving a backward Hamilton-Jacobi equation whereas the distribution of the agents’ states evolves according to a forward Fokker-Planck equation. The “Master" equation (stated in the space of probability measures) subsumes the individual and collective behaviors. Finally, modeling, numerical analysis and scientific computing are crucial for the applications. French mathematicians play a world-leading role in the research on MFG: The terminology itself comes from a series of pioneering works by J.-M. Lasry and P.-L. Lions who introduced most of the key ideas for the mathematical analysis of MFG; the last conference on MFG was held last June in Paris and organized by Y. Achdou, P. Cardaliaguet and J.-M. Lasry. As testifies the proposal, the number of researchers working on MFG in France (and also abroad) is extremely fast-growing, not only because the theoretical aspects are exciting and challenging, but also because MFG models find more and more applications. The aim of the project is to better coordinate the French mathematical research on MFG and to achieve significant progress in the theory and its applications.

The partners of the project are the CEREMADE laboratory (Paris Dauphine), the IRMAR laboratory (Rennes I), the university of Nice and of Tours.

5.1.3. IPL (FRATRES)

IPSO is associated to IPL FRATRES which started in june 2015. The aim of this project is to organize Inria teams activities which develop mathematical and numerical tools in magnetically confined nuclear fusion. The ambition is to prepare the next generation of numerical modeling methodologies able to use in an optimal way the processing capabilities of modern massively parallel architectures. This objective requires close collaboration between a) applied mathematicians and physicists that develop and study mathematical models of PDE; b) numerical analysts developing approximation schemes; c) specialists of algorithmics proposing solvers and libraries using the many levels of parallelism offered by the modern architecture and d) computer scientists. The project road map ambitions to contribute in close connection with National and European initiatives devoted to nuclear Fusion to the improvement and design of numerical simulation technologies applied to plasma physics and in particular to the ITER project for magnetic confinement fusion.

Postdoc

- Xiaofei Zhao has been hired as a Postdoc, under the supervision of Nicolas Crouseilles and Sever Hirstoaga (Inria-Nancy). His contract started in october 2015 and will end in august 2016.

5.2. European Initiatives

5.2.1. FP7 & H2020 Projects

Project acronym: GEOPARDI

Program: FP7

Project title: Numerical integration of Geometric Partial Differential Equations

Duration: September 2011 - August 2016

Coordinator: Erwan Faou, Inria

Abstract: The goal of this project is to develop new numerical methods for the approximation of evolution equations possessing strong geometric properties such as Hamiltonian systems or stochastic differential equations. In such situations the exact solutions endow with many physical properties that are consequences of the geometric structure: Preservation of the total energy, momentum conservation or existence of ergodic invariant measures. However the preservation of such qualitative properties of the original system by numerical methods at a reasonable cost is not guaranteed at all, even for very precise (high order) methods. The principal aim of geometric numerical integration is the understanding and analysis of such problems: How (and to which extend) reproduce qualitative behavior of differential equations over long time? The extension of this theory to partial differential equations is a fundamental ongoing challenge, which require the invention of a new mathematical framework bridging the most recent techniques used in the theory of nonlinear PDEs and stochastic ordinary and partial differential equations. The development of new efficient numerical schemes for geometric PDEs has to go together with the most recent progress in analysis (stability phenomena, energy transfers, multiscale problems, etc..) The major challenges of the project are to derive new schemes by bridging the world of numerical simulation and the analysis community, and to consider deterministic and stochastic equations, with a general aim at deriving hybrid methods. We also aim to create a research platform devoted to extensive numerical simulations of difficult academic PDEs in order to highlight new nonlinear phenomena and test numerical methods.

Erwan Faou was the principal investigator of the ERC Starting Grant Project Geopardi (2011-2016).

Between 2011 and 2016, Erwan Faou was the principal investigator of this ERC Starting grant project. This research project is centered on the numerical simulation of geometric evolution partial differential equations (PDEs). Typical examples are given by Hamiltonian Partial Differential Equations (PDE) such as wave equations in nonlinear propagations problems, Schrödinger equations in quantum mechanics, or Vlasov equations in plasma physics. The main goals of the project can be summarized as follows:

- Analyze numerical schemes for Hamiltonian PDEs and stochastic differential equations as mathematical objects in their own right, and study their global behavior (invariant preservation, ergodicity with respect to some invariant measure, averaging properties, scattering, etc...)
- Develop new numerical methods in connection with the most recent advances in the theoretical studies, and devoted to specific situations (high frequency computations, stochastic and hybrid methods, Vlasov and Euler equations). In particular, an important objective is the analysis of the long time behavior of these equations.

The main originality of the Geopardi project is the combination of rigorous nonlinear analysis, numerical analysis and numerical simulations, as well as its hybrid nature mixing deterministic and stochastic problems. The project has an excellent international visibility. The participants have been invited in many conferences to present their works in the last year (Scicade 13 & 15, Numdiff 13, workhops in Toronto, Harvard, IHES, Oberwolfach or Luminy, etc..). The research outcomes are published in high level international journals such as J. Amer. Math. Soc., Numer. Math., SIAM J. Numer. Anal. or Math. Comp. The project has also been used to invite collaborators and researcher to visit Inria. In particular, E. Faou organized with T. Lelièvre and J. Erhel in september 2013 the NASPDE conference whose main topic is the numerical simulation of stochastic PDEs, and that was mainly funded by the Geopardi project.

5.2.2. Collaborations in European Programs, Except FP7 & H2020

Project acronym: WPENR

Program: EUROFusion Enabling Research project ER15-IPP-01

Project title: Verification and development of new algorithms for gyrokinetic codes

Duration: January 2015 - December 2018

Coordinator: Eric Sonnenndrücker (Max-Planck-Institut für Plasmaphysik (IPP), Germany)

Other partners: IPP (Germany), EPFL (Switzerland), CEA-Cadarache (France), university of Strasbourg, Toulouse, Marseille, Paris 6 (France).

Abstract: Gyrokinetic codes play a major role in understanding the development and saturation of micro- turbulence in a magnetic fusion plasma and its influence on energy confinement time. The first aim of this proposal is to assess the reliability of gyrokinetic codes by extensive verification and benchmarking. All the major european gyrokinetic codes are involved in the proposal and this will enable them to define comparison elements, which ultimately will also facilitate the cross-validation of new physics. On the other hand we will develop new algorithms for extending the physics capabilities or the computational efficiency of different gyrokinetic codes. Finally we will also perform a prospective investigation of models and numerical methods that could help in the future to address physics where kinetic effects might play an important role but that cannot be handled with today's gyrokinetic codes, like L-H (low to high confinement) transition, edge physics or MHD time scales simulations.

5.3. International Research Visitors

5.3.1. Visits of International Scientists

- Philippe Chartier and Nicolas Crouseilles invited Eric Sonnendrücker (IPP Max Planck) for one week in june 2016.
- Nicolas Crouseilles and Mohammed Lemou invited Shi Jin and Liu Liu (university of Wisconsin) for two weeks in june 2016.
- Arnaud Debussche invited Martina Hofmanova (TU Berlin) for one week in november 2016.
- Erwan Faou invited Chuchu Chen (Michigan state university) for two weeks in november 2016.

5.3.2. Visits to International Teams

5.3.2.1. Research Stays Abroad

- Philippe Chartier was invited for a one-week working visit by Gilles Vilmart, university of Geneva (Switzerland).
- Nicolas Crouseilles was invited for a one-week working visit by Gilles Vilmart, university of Geneva (Switzerland).
- Arnaud Debussche was invited at SNS Pisa (Italy) for two periods of one week in april and november 2016.
- Erwan Faou was invited in the university of Trondheim (Norway) in october 2016.

6. Dissemination

6.1. Promoting Scientific Activities

6.1.1. Scientific Events Organisation

6.1.1.1. Member of the Organizing Committees

- François Castella and Philippe Chartier organized the workshop "Multiscale methods for Schödinger and kinetic equations", Saint-Malo (France), december 12-14, 2016.
- Arnaud Debussche organized the conference "Stochastic Partial Differential Equations and Applications-X, Levico Terme (Italy), may 30-june 4, 2016.
- Erwan Faou organized the workshop "Geometric Numerical Integration", Oberwolfach (Germany), march 20-26, 2016. Co-organized with E. Hairer, M. Hochbruck and C. Lubich.

6.1.2. Journal

6.1.2.1. Member of the Editorial Boards

- Philippe Chartier is member of the editorial board of "Mathematical Modelling and Numerical Analysis" (2007-).
- Arnaud Debussche is editor in chief of the journal "Stochastics and Partial Differential Equations: analysis and computations".
- Arnaud Debussche is member of the editorial board of Potential Analysis (2011-).
- Arnaud Debussche is member of the editorial board of Differential and Integral Equations (2002-).
- Arnaud Debussche is member of the editorial board of ESAIM:PROC (2012-).
- Arnaud Debussche is member of the editorial board of Journal of Evolution Equation (2014-).
- Arnaud Debussche is member of the editorial board of Applied Mathematics & Optimization (2014-).
- Arnaud Debussche is member of the editorial board of the collection : "Mathématiques & Applications" (Springer).
- Erwan Faou was editor of the Oberwolfach reports [31] (2016).

6.1.2.2. Reviewer - Reviewing Activities

Members of IPSO are reviewers for almost the journals in which they publish.

6.1.3. Invited Talks

- Philippe Chartier was invited speaker at the workshop "Mould calculus, from multiple zeta values to B-series", Pau (France), december 1-2, 2016.
- Philippe Chartier was plenary speaker at the international conference ICNAAM, Rhodes (Greece), september 2016.
- Philippe Chartier was invited speaker at the workshop "GAMPP", IPP Garching (Germany), september 12-16, 2016.
- Philippe Chartier was invited speaker at the workshop "Stability and discretization issues in differential equations", Trieste (Italy), june 2016.
- Philippe Chartier gave a seminar at the university of Lille (France), june 9, 2016.
- Philippe Chartier was invited speaker at Meeting ANR Moonrise, Toulouse (France), june 2-3, 2016.
- Philippe Chartier gave a seminar at the university of Geneva (Switzerland), may 26-june 1, 2016.
- Philippe Chartier was invited at the workshop "Geometric Numerical Integration", Oberwolfach (Germany), march 20-26, 2016.
- Nicolas Crouseilles was invited at the workshop "Geometric Numerical Integration", Oberwolfach (Germany), march 20-26, 2016.
- Nicolas Crouseilles gave a seminar at the university of Geneva (Switzerland), may 13, 2016.
- Nicolas Crouseilles gave a seminar at the university of Paris Sud, Orsay (France), november 17, 2016.
- Nicolas Crouseilles was invited speaker at the workshop "NumKin", Strasbourg (France), october 17-21, 2016.
- Nicolas Crouseilles was invited speaker at the workshop "Kinet", Madison (US), april 21-25, 2016.
- Arnaud Debussche was invited speaker at the workshop "Probabilistic models-from discrete to continuous", university of Warwick (UK), march 29-april 2, 2016.
- Arnaud Debussche was invited speaker at the workshop "Stochastic Analysis and Related Fields", Humboldt university Berlin (Germany), july 28-30, 2016.

- Arnaud Debussche was invited speaker at the workshop "Nonlinear Wave and Dispersive Equations", Kyoto university (Japan), september 6-8, 2016.
- Arnaud Debussche was invited speaker at the workshop "Nonlinear Stochastic Evolution Equations: Analysis and Numerics", TU Berlin (Germany), november 3-5, 2016.
- Erwan Faou gave a seminar at the CERMICS, Marne-La-Vallée (France), december 2016.
- Erwan Faou was invited at the workshop "Structure and scaling in computational field theories", Oslo (Norway), november 2016.
- Erwan Faou was invited at the conference "Nonlinear waves", IHES (France), may 2016.
- Erwan Faou was invited at the workshop "Nonlinear Evolution Problems", Oberwolfach (Germany), march 2016.
- Erwan Faou was invited at the workshop "Recent trends in nonlinear evolution equations", CIRM-Luminy (France), april 4-8, 2016.
- Mohammed Lemou was plenary speaker at the workshop "Asymptotic behavior of systems of PDE arising in physics and biology: theoretical and numerical points of view", Lille (France), june 2016.
- Mohammed Lemou was invited speaker at the workshop "NumKin", Strasbourg (France), october 17-21, 2016.
- Mohammed Lemou was invited speaker at the workshop "Kinet", Madison (US), april 21-25, 2016.
- Mohammed Lemou was invited speaker at the ANR Moonrise Meeting, Toulouse (France), june 2016.
- Florian Méhats was plenary speaker at the workshop "Journée des jeunes EDPistes français", Bordeaux (France).
- Florian Méhats gave a seminar of the university of Paris Sud, Orsay (France).
- Florian Méhats gave a seminar of the university of Nice (France).
- Florian Méhats gave a seminar of the university of Lille (France).

6.1.4. Scientific Expertise

- Philippe Chartier was member of the hiring committee of an associate professor, university of Trondheim (Norway).
- Philippe Chartier was member of the hiring committee CR2-Inria (Bordeaux).
- Nicolas Crouseilles was member of the CORDI-S committee at Inria-Rennes.
- Arnaud Debussche was a member of the hiring committee of a professor, university of Rennes 1
- Arnaud Debussche was a member of the hiring committee of a "Maître de conférence", university of Orléans.
- Mohammed Lemou was member of the hiring committee of a professor, university of Rennes 1.
- Mohammed Lemou was was a member of the hiring committee of a "Maître de conférence", university of Nantes.

6.1.5. Research Administration

- François Castella is member of the IRMAR laboratory council.
- Philippe Chartier is the vice-head of science (DSA) of the Rennes Inria-Center.
- Philippe Chartier is member of the direction committee (ED) of the Rennes Inria-Center.
- Philippe Chartier is member of the national evaluation committee (CE) of Inria.
- Nicolas Crouseilles is member of the Scientific Council of the ENS Rennes.
- Nicolas Crouseilles is member of the committee of the Fédération de Fusion".
- Arnaud Debussche is vice president in charge of research and international relations of the Ecole Normale Supérieure de Rennes.

- Arnaud Debussche is member of the executive board of the Lebesgue Center.
- Arnaud Debussche is director of the "Agence Lebesgue de Mathématiques pour l'Innovation".
- Erwan Faou was member of the COST-GTRI (Comité d'orientation scientifique et technologique, groupe de travail pour les relations internationales) at Inria.
- Erwan Faou is member of the Scientific Council of the Pôle Universitaire Léonard de Vinci.
- Erwan Faou is member of the CNU 26.
- Mohammed Lemou is member of the Scientific Council of the ENS Rennes.
- Mohammed Lemou is member of the Scientific Council of the Lebesgue Center.
- Mohammed Lemou is head of the team "analyse numérique" of IRMAR laboratory.
- Florian Méhats is head of the IRMAR laboratory.

6.2. Teaching - Supervision - Juries

6.2.1. Teaching

- François Castella gave a course in M1 on kinetic equations, university of Rennes 1 (60 hours).
- Philippe Chartier gave a course in L3 on ordinary differential equations, Ecole Normale Supérieure de Rennes (24 hours).
- Philippe Chartier gave a course in M2 on geometric numerical integration and averaging methods, university of Rennes 1 (24 hours).
- Nicolas Crouseilles gave a course in M2 on numerical methods for kinetic equations, university of Rennes 1 (12 hours).
- Arnaud Debussche gave a course in M2 on stochastic partial differential equations, university of Rennes 1 (24 hours).
- Erwan Faou gave a course in M1 on modelisation and numerical analysis of PDEs, ENS Paris, in collaboration with E. Dormy.
- Mohammed Lemou gave a course in M2 on partial differential equations, university of Rennes 1 (24 hours).
- Mohammed Lemou is head of the M2 "Analyse et Applications".

6.2.2. Supervision

- François Castella supervises the PhD thesis of Valentin Doli, *Mathematical and ecological study of the propagation of a specific virus attacking plants*, (2014-). Co-advisor: Frédéric Hamelin (Agro-Rennes).
- François Castella and Philippe Chartier supervised the PhD thesis of Julie Sauzeau, *Highly-oscillatory central manifold and application to ecology* (2013-2016). Julie Sauzeau is now teacher.
- Nicolas Crouseilles and Erwan Faou supervise the PhD thesis of Joackim Bernier, *Mathematical and numerical anaysis of nonlinear transport equations*, (2016-).
- Nicolas Crouseilles and Mohammed Lemou supervised the PhD thesis of Hélène Hivert *Mathematical and numerical study of kinetic model and their asymptotics: diffusion and anomalous diffusion limit*, (2013-2016). Hélène Hivert is now post-doc at ENS Lyon.
- Erwan Faou supervises the PhD thesis of Romain Horsin, *Mathematical and numerical analysis of the Vlasov-HMF model*, (2014-). Co-advisor: Frédéric Rousset (university Paris Sud Orsay).
- Arnaud Debussche supervizes the PhD thesis of Mac Jugal Nankep *PDMP with spatial dependency for the dynamics of gene networks*, (2014-).
- Arnaud Debussche and Florian Méhats are supervisors of the PhD thesis of Maxime Tusseau. *Highly oscillatory nonlinear Schrödinger equation with stochastic potential*, (2013-).

- Mohammed Lemou and Florian Méhats are supervisors of the PhD thesis of Marine Malo *Collisionless kinetic equations: stability, oscillations*, (2015-).

6.2.3. Juries

- Erwan Faou was referee of the PhD thesis of Ahmed-Amine Homman (CEA and ENPC), june 2016.
- Nicolas Crouseilles was referee of the PhD thesis of Mehdi Badsi (university Paris 6), october 2016.
- Nicolas Crouseilles was referee of the PhD thesis of Nhung Pham (university of Strasbourg), december 2016.
- Nicolas Crouseilles was member of the jury of the PhD thesis of Julie Sauzeau (university of Rennes 1), june 2016.
- Arnaud Debussche was referee of the PhD thesis of Nathalie Ayi (university of Nice), june 2016.
- Arnaud Debussche was member of jury of the PhD thesis of Vincent Renault (university of Paris 6), september 2016.
- Mohammed Lemou was referee of the PhD thesis of Thomas Le Roy (university Paris 6), january 2016.
- Mohammed Lemou was referee of the PhD thesis of Ankit Ruhi (IIS, Bangalore, India), december 2016.
- Mohammed Lemou was member of the jury of the PhD thesis of Sébastien Guisset (university of Bordeaux 1), september 2016.

7. Bibliography

Publications of the year

Articles in International Peer-Reviewed Journal

[1] S. BALAC, A. FERNANDEZ, F. MAHÉ, F. MÉHATS, R. TEXIER-PICARD.*The Interaction Picture method for solving the generalized nonlinear Schrödinger equation in optics*, in "ESAIM: Mathematical Modelling and Numerical Analysis", June 2016, vol. 50, n^o^ 4, p. 945-964 [*DOI :* 10.1051/M2AN/2015060], https://hal.archives-ouvertes.fr/hal-00850518.

[2] V. BANICA, E. FAOU, E. MIOT.*Collision of almost parallel vortex filaments*, in "Communications on Pure and Applied Mathematics", 2017, vol. 70, n^o^ 2, p. 378-405 [*DOI :* 10.1002/CPA.21637], https://hal.archives-ouvertes.fr/hal-01170929.

[3] W. BAO, L. LE TREUST, F. MÉHATS.*Dimension reduction for dipolar Bose-Einstein condensates in the strong interaction regime*, in "Kinetic and Related Models ", September 2017, https://hal.archives-ouvertes.fr/hal-01101793.

[4] A. BOUILLARD, E. FAOU, M. ZAVIDOVIQUE.*Fast Weak-Kam Integrators for separable Hamiltonian systems*, in "Mathematics of Computation", 2016, vol. 85, n^o^ 297, p. 85-117, https://hal.archives-ouvertes.fr/hal-00743462.

[5] F. CASAS, N. CROUSEILLES, E. FAOU, M. MEHRENBERGER.*High-order Hamiltonian splitting for Vlasov-Poisson equations*, in "Numerische Mathematik", 2016, https://hal.inria.fr/hal-01206164.

[6] F. CASTELLA, S. MADEC, Y. LAGADEUC.*Global behavior of N competing species with strong diffusion: diffusion leads to exclusion*, in "Applicable Analysis", 2016, vol. 95, n^o^ 2, p. 341-372 [*DOI :* 10.1080/00036811.2015.1004320], https://hal.archives-ouvertes.fr/hal-01026195.

[7] P. CHARTIER, N. J. MAUSER, F. MÉHATS, Y. ZHANG.*Solving highly-oscillatory NLS with SAM: numerical efficiency and geometric properties*, in "Discrete and Continuous Dynamical Systems - Series S", October 2016, vol. 9, n^{o} 5, p. 1327-1349 [*DOI :* 10.3934/DCDSS.2016053], https://hal.archives-ouvertes.fr/hal-00850513.

[8] P. CHARTIER, F. MÉHATS, M. THALHAMMER, Y. ZHANG.*Improved error estimates for splitting methods applied to highly-oscillatory nonlinear Schrödinger equations*, in "Mathematics of Computation", 2016, vol. 85, n^{o} 302, p. 2863-2885 [*DOI :* 10.1090/MCOM/3088], https://hal.archives-ouvertes.fr/hal-01373280.

[9] A. CHERNOV, A. DEBUSSCHE, F. NOBILE.*Numerical methods for random and stochastic partial differential equations*, in "Stochastics and Partial Differential Equations Analysis and Computations", March 2016, vol. 4, n^{o} 1, p. 1-2 [*DOI :* 10.1007/S40072-016-0073-2], https://hal.archives-ouvertes.fr/hal-01297401.

[10] P. J. COELHO, N. CROUSEILLES, P. PEREIRA, M. ROGER.*Multi-scale methods for the solution of the radiative transfer equation*, in "Journal of Quantitative Spectroscopy and Radiative Transfer", March 2016, vol. 172, p. 36-49 [*DOI :* 10.1016/J.JQSRT.2015.10.001], https://hal.archives-ouvertes.fr/hal-01274991.

[11] N. CROUSEILLES, G. DIMARCO, M. LEMOU.*Asymptotic preserving and time diminishing schemes for rarefied gas dynamic*, in "Kinetic and Related Models ", 2016, https://hal.inria.fr/hal-01392412.

[12] N. CROUSEILLES, G. DIMARCO, M.-H. VIGNAL.*Multiscale schemes for the BGK-Vlasov-Poisson system in the quasi-neutral and fluid limits. Stability analysis and first order schemes*, in "Multiscale Modeling and Simulation: A SIAM Interdisciplinary Journal", 2016, vol. 14, n^{o} 1, p. 65-95 [*DOI :* 10.1137/140991558], https://hal.inria.fr/hal-01090676.

[13] N. CROUSEILLES, L. EINKEMMER, E. FAOU.*An asymptotic preserving scheme for the relativistic Vlasov–Maxwell equations in the classical limit*, in "Computer Physics Communications", December 2016, vol. 209, https://hal.inria.fr/hal-01283779.

[14] N. CROUSEILLES, H. HIVERT, M. LEMOU.*Numerical schemes for kinetic equations in the anomalous diffusion limit. Part II: degenerate collision frequency.* , in "SIAM Journal on Scientific Computing", 2016, vol. 38, n^{o} 4, https://hal.inria.fr/hal-01245312.

[15] N. CROUSEILLES, H. HIVERT, M. LEMOU.*Numerical schemes for kinetic equations in the diffusion and anomalous diffusion limits. Part I: the case of heavy-tailed equilibrium*, in "SIAM J. Sci. Comput.", 2016, vol. 38, n^{o} 2, p. 737-764, https://hal.inria.fr/hal-01130002.

[16] N. CROUSEILLES, M. LEMOU, F. MÉHATS, X. ZHAO.*Uniformly accurate forward semi-Lagrangian methods for highly oscillatory Vlasov-Poisson equations*, in "Multiscale Modeling and Simulation: A SIAM Interdisciplinary Journal", 2017, https://hal.inria.fr/hal-01286947.

[17] N. CROUSEILLES, M. LEMOU, S. RAGHURAMA RAO, A. RUHI, M. SEKHAR.*Asymptotic Preserving scheme for a kinetic model describing incompressible fluids*, in "Kinetic and Related Models ", 2016, vol. 9, p. 51-74, https://hal.inria.fr/hal-01090677.

[18] N. CROUSEILLES, M. LEMOU, G. VILMART.*Asymptotic Preserving numerical schemes for multiscale parabolic problems*, in "Comptes Rendus Mathématique", 2016, vol. 354, n^{o} 3, p. 271-276, 7 pages [*DOI :* 10.1016/J.CRMA.2015.11.010], https://hal.archives-ouvertes.fr/hal-01187092.

[19] G. DA PRATO, A. DEBUSSCHE.*An integral inequality for the invariant measure of a stochastic reaction–diffusion equation*, in "Journal of Evolution Equations", 2016, p. 1-18 [*DOI :* 10.1007/s00028-016-0349-z], https://hal.archives-ouvertes.fr/hal-01235038.

[20] G. DA PRATO, A. DEBUSSCHE.*Estimate for P_tD for the stochastic Burgers equation*, in "Annales de l'Institut Henri Poincaré (B) Probabilités et Statistiques", 2016, vol. 52, n^{o} 3, p. 1248-1258 [*DOI :* 10.1214/15-AIHP685], https://hal.archives-ouvertes.fr/hal-01110454.

[21] A. DEBUSSCHE, S. DE MOOR, J. VOVELLE.*Diffusion limit for the radiative transfer equation perturbed by a Markovian process*, in "Asymptotic Analysis", 2016, vol. 98, n^{o} 1-2, p. 31-58, 24 pages. arXiv admin note: text overlap with arXiv:1311.4743 [*DOI :* 10.3233/ASY-161360], https://hal.archives-ouvertes.fr/hal-01044426.

[22] A. DEBUSSCHE, M. HOFMANOVÁ, J. VOVELLE.*Degenerate Parabolic Stochastic Partial Differential Equations: Quasilinear case*, in "The Annals of Probability", 2016, vol. 44, n^{o} 3, p. 1916-1955 [*DOI :* 10.1214/15-AOP1013], https://hal.archives-ouvertes.fr/hal-00863829.

[23] E. FAOU, P. GERMAIN, Z. HANI.*The weakly nonlinear large-box limit of the 2D cubic nonlinear Schrödinger equation*, in "Journal of the American Mathematical Society", 2016, vol. 29, n^{o} 4, p. 915-982, 68 pages, 1 figure [*DOI :* 10.1090/JAMS/845], https://hal.archives-ouvertes.fr/hal-00905851.

[24] E. FAOU, R. HORSIN, F. ROUSSET.*On numerical Landau damping for splitting methods applied to the Vlasov-HMF model*, in "Foundations of Computational Mathematics", 2016, 38 [*DOI :* 10.1007/s10208-016-9333-9], https://hal.archives-ouvertes.fr/hal-01219115.

[25] E. FAOU, F. ROUSSET.*Landau damping in Sobolev spaces for the Vlasov-HMF model*, in "Archive for Rational Mechanics and Analysis", 2016, vol. 219, n^{o} 2, p. 887-902 [*DOI :* 10.1007/s00205-015-0911-9], https://hal.archives-ouvertes.fr/hal-00956595.

[26] S. FOURNAIS, L. LE TREUST, N. RAYMOND, J. VAN SCHAFTINGEN.*Semiclassical Sobolev constants for the electro-magnetic Robin Laplacian*, in "Journal of the Mathematical Society of Japan", 2017, https://hal.archives-ouvertes.fr/hal-01285311.

[27] F. M. HAMELIN, F. CASTELLA, V. DOLI, B. MARÇAIS, V. RAVIGNÉ, M. A. LEWIS.*Mate Finding, Sexual Spore Production, and the Spread of Fungal Plant Parasites*, in "Bulletin of Mathematical Biology", April 2016, vol. 78, n^{o} 4, p. 695-712 [*DOI :* 10.1007/s11538-016-0157-1], https://hal.archives-ouvertes.fr/hal-01313012.

[28] M. LEMOU.*Extended Rearrangement inequalities and applications to some quantitative stability results*, in "Communications in Mathematical Physics", 2016, vol. 348, n^{o} 2, p. 695-727 [*DOI :* 10.1007/s00220-016-2750-4], https://hal.archives-ouvertes.fr/hal-01207621.

[29] F. MÉHATS, C. SPARBER.*Dimension reduction for rotating Bose-Einstein condensates with anisotropic confinement*, in "Discrete and Continuous Dynamical Systems - Series A", 2016, vol. 36, n^{o} 9, p. 5097-5118, 22 pages [*DOI :* 10.3934/DCDS.2016021], https://hal.archives-ouvertes.fr/hal-01178431.

Scientific Books (or Scientific Book chapters)

[30] M. CHAUSSADE-BEAUDOUIN, M. DAUGE, E. FAOU, Z. YOSIBASH.*High frequency oscillations of first eigenmodes in axisymmetric shells as the thickness tends to zero*, in "Operator Theory Advances and Application", Recent Trends in Operator Theory and Partial Differential Equations - The Roland Duduchava Anniversary Volume, Birkhäuser/Springer, 2017, vol. 258, p. 89-110, https://hal.archives-ouvertes.fr/hal-01278861.

Research Reports

[31] E. FAOU, E. HAIRER, M. HOCHBRUCK, C. LUBICH.*Geometric Numerical Integration*, Mathematisches Forschungsinstitut Oberwolfach, 2016, n° 18/2016, p. 869 - 948, Oberwolfach Reports, vol. 13 n° 1 (2016) [*DOI :* 10.4171/OWR/2016/18], https://hal.archives-ouvertes.fr/hal-01403326.

Other Publications

[32] N. ARRIZABALAGA, L. LE TREUST, N. RAYMOND.*On the MIT bag model: self-adjointness and non-relativistic limit*, July 2016, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01343717.

[33] I. BAILLEUL, A. DEBUSSCHE, M. HOFMANOVA.*Quasilinear generalized parabolic Anderson model*, 2016, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01389489.

[34] S. BAUMSTARK, E. FAOU, K. SCHRATZ.*Uniformly accurate exponential-type integrators for klein-gordon equations with asymptotic convergence to classical splitting schemes in the nonlinear schroedinger limit*, June 2016, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01331949.

[35] F. CASTELLA, P. CHARTIER, J. SAUZEAU.*A formal series approach to the center manifold theorem*, April 2016, working paper or preprint, https://hal.inria.fr/hal-01279715.

[36] S. CERRAI, A. DEBUSSCHE.*Large deviations for the two-dimensional stochastic Navier-Stokes equation with vanishing noise correlation*, 2016, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01287049.

[37] P. CHARTIER, N. CROUSEILLES, M. LEMOU.*An averaging technique for transport equations*, September 2016, working paper or preprint, https://hal.inria.fr/hal-01374655.

[38] P. CHARTIER, N. CROUSEILLES, M. LEMOU.*Averaging of highly-oscillatory transport equations*, November 2016, working paper or preprint, https://hal.inria.fr/hal-01396685.

[39] P. CHARTIER, L. LE TREUST, F. MÉHATS.*Uniformly accurate time-splitting methods for the semiclassical Schrödinger equationPart 2 : Numerical analysis of the linear case*, January 2016, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01257753.

[40] P. CHARTIER, M. LEMOU, F. MÉHATS.*Highly-oscillatory evolution equations with multiple frequencies: averaging and numerics*, October 2016, working paper or preprint, https://hal.inria.fr/hal-01281950.

[41] M. CHAUSSADE-BEAUDOUIN, M. DAUGE, E. FAOU, Z. YOSIBASH.*Free Vibrations of Axisymmetric Shells: Parabolic and Elliptic cases*, January 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01264125.

[42] N. CROUSEILLES, S. JIN, M. LEMOU.*Nonlinear Geometric Optics method based multi-scale numerical schemes for highly-oscillatory transport equations*, May 2016, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01323721.

[43] N. CROUSEILLES, M. LEMOU, F. MÉHATS, X. ZHAO.*Uniformly accurate Particle-In-Cell method for the long time two-dimensional Vlasov-Poisson equation with strong magnetic field*, December 2016, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01418976.

[44] A. DEBUSSCHE, H. WEBER.*The Schrödinger equation with spatial white noise potential*, 2016, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01411205.

[45] E. FAOU, T. JÉZÉQUEL.*Convergence of a normalized gradient algorithm for computing ground states*, March 2016, working paper or preprint, https://hal.inria.fr/hal-01284679.

[46] R. L. FRANK, F. MÉHATS, C. SPARBER.*Averaging of nonlinear Schrödinger equations with strong magnetic confinement*, 2016, 12 pages, https://hal.archives-ouvertes.fr/hal-01397325.

[47] M. LEMOU, F. MÉHATS, X. ZHAO.*Uniformly accurate numerical schemes for the nonlinear dirac equation in the nonrelativistic limit regime*, May 2016, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01313976.

References in notes

[48] E. HAIRER.*Geometric integration of ordinary differential equations on manifolds*, in "BIT", 2001, vol. 41, p. 996–1007.

[49] E. HAIRER, C. LUBICH, G. WANNER.*Geometric Numerical Integration. Structure-Preserving Algorithms for Ordinary Differential Equations, Second edition*, Springer Series in Computational Mathematics 31, Springer, Berlin, 2006.

[50] E. HAIRER, G. WANNER.*Solving Ordinary Differential Equations II. Stiff and Differential-Algebraic Problems*, Springer Series in Computational Mathematics 14, 2, Springer-Verlag, Berlin, 1996.

[51] C. LUBICH.*A variational splitting integrator for quantum molecular dynamics*, in "Appl. Numer. Math.", 2004, vol. 48, p. 355–368.

[52] C. LUBICH.*On variational approximations in quantum molecular dynamics*, in "Mathematics of Computation", 2009.

[53] J. M. SANZ-SERNA, M. P. CALVO.*Numerical Hamiltonian Problems*, Chapman & Hall, London, 1994.

Project-Team KERDATA

Scalable Storage for Clouds and Beyond

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

IN PARTNERSHIP WITH:
Institut national des sciences appliquées de Rennes

Université Rennes 1

École normale supérieure de Rennes

RESEARCH CENTER
Rennes - Bretagne-Atlantique

THEME
Distributed and High Performance Computing

Table of contents

Project-Team KERDATA

Creation of the Team: 2009 July 01, updated into Project-Team: 2012 July 01

Keywords:

Computer Science and Digital Science:

1.1.4. - High performance computing
1.1.5. - Exascale
1.1.6. - Cloud
1.3. - Distributed Systems
1.6. - Green Computing
2.6.2. - Middleware
3.1.3. - Distributed data
3.1.8. - Big data (production, storage, transfer)
3.3.3. - Big data analysis
6.2.7. - High performance computing
7.1. - Parallel and distributed algorithms

Other Research Topics and Application Domains:

1.1.2. - Molecular biology
2.6.1. - Brain imaging
3.2. - Climate and meteorology
4.5.1. - Green computing
9.4.5. - Data science

1. Members

Research Scientists

Gabriel Antoniu [Team leader, Inria, Senior Researcher, HDR]
Shadi Ibrahim [Inria, Researcher]

Faculty Members

Luc Bougé [ENS Rennes, Professor, HDR]
Alexandru Costan [INSA Rennes, Associate Professor]

Technical Staff

Hadi Salimi [Inria, from April 2016]

PhD Students

Nathanaël Cheriere [ENS Rennes, from September 2016]
Paul Le Noac'h [INSA Rennes, from November 2016]
Ovidiu-Cristian Marcu [Inria]
Pierre Matri [Inria and Universidad Politécnica de Madrid]
Tien-Dat Phan [Univ. Rennes I]
Luis Eduardo Pineda Morales [Inria]
Lokman Rahmani [Univ. Rennes I]
Mohammed-Yacine Taleb [Inria]
Orçun Yildiz [Inria]

Post-Doctoral Fellow

Chi Zhou [Inria, from May 2016]

Administrative Assistant

Aurélie Patier [Univ. Rennes I]

Others

Muhammad Najeeb Aslam [Univ. Rennes I, Master intern, from May 2016 until August 2016]

Rémi Hutin [Inria, Undergraduate intern, from May 2016 until July 2016]

2. Overall Objectives

2.1. Context: the need for scalable data management

We are witnessing a rapidly increasing number of application areas generating and processing very large volumes of data on a regular basis. Such applications are called *data-intensive*. Governmental and commercial statistics, climate modeling, cosmology, genetics, bio-informatics, high-energy physics are just a few examples in the scientific area. In addition, rapidly growing amounts of data from social networks and commercial applications are now routinely processed.

In all these examples, the overall application performance is highly dependent on the properties of the underlying data management service. It becomes crucial to store and manipulate massive data efficiently. However, these data are typically *shared* at a large scale and *concurrently accessed* at a high degree. With the emergence of recent infrastructures such as cloud computing platforms and post-Petascale high-performance computing (HPC) systems, achieving highly scalable data management under such conditions has become a major challenge.

2.1.1. Our objective

The KerData project-team is namely focusing on designing innovative architectures and systems for *scalable data storage and processing*. We target two types of infrastructures: *clouds* and *post-Petascale high-performance supercomputers*, according to the current needs and requirements of data-intensive applications.

We are especially concerned by the applications of major international and industrial players in cloud computing and extreme-scale high-performance computing (HPC), which shape the long-term agenda of the cloud computing [40], [37] and Exascale HPC [39] research communities. The Big Data area, which has recently captured a lot of attention, emphasized the challenges related to Volume, Velocity and Variety. This is yet another element of context that further highlights the primary importance of designing data management systems that are efficient at a very large scale.

2.1.1.1. Alignment with Inria's scientific strategy

Data-intensive applications exhibit several common requirements with respect to the need for data storage and I/O processing. We focus on some core challenges related to data management, resulted from these requirements. Our choice is perfectly in line with Inria's strategic plan [44], which acknowledges as critical the challenges of *storing, exchanging, organizing, utilizing, handling and analyzing* the huge volumes of data generated by an increasing number of sources. This topic is also stated as a scientific priority of Inria's research centre of Rennes [43]: *Storage and utilization of distributed big data.*

2.1.1.2. Challenges and goals related to cloud data storage and processing

In the area of cloud data processing, a significant milestone is the emergence of the Map-Reduce [50] parallel programming paradigm. It is currently used on most cloud platforms, following the trend set up by Amazon [35]. At the core of Map-Reduce frameworks lies the storage system, a key component which must meet a series of specific requirements that are not fully met yet by existing solutions: the ability to provide efficient *fine-grain access* to the files, while sustaining a *high throughput* in spite of *heavy access concurrency*; the need to provide a high resilience to *failures*; the need to take *energy-efficiency* issues into account.

More recently, it becomes clear that data-intensive processing needs to go beyond the frontiers of single datacenters. In this perspective, extra challenges arise, related to the efficiency of metadata management. This efficiency has a major impact on the access to very large sets of small objects by Big Data processing workflows running on large-scale infrastructures.

2.1.1.3. Challenges and goals related to data-intensive HPC applications

Key research fields such as climate modeling, solid Earth sciences or astrophysics rely on very large-scale simulations running on post-Petascale supercomputers. Such applications exhibit requirements clearly identified by international panels of experts like IESP [42], EESI [38], ETP4HPC [39]. A jump of one order of magnitude in the size of numerical simulations is required to address some of the fundamental questions in several communities in this context. In particular, the lack of data-intensive infrastructures and methodologies to analyze the huge results of such simulations is a major limiting factor.

The challenge we have been addressing is to find new ways to store, visualize and analyze massive outputs of data during and after the simulations. Our main initial goal was to do it without impacting the overall performance, avoiding the *jitter* generated by I/O interference as much as possible. Recently, we started to focus specifically on *in situ processing* approaches and we explored approaches to *model and predict I/O phase occurrences* and to *reduce intra-application and cross-application I/O interference*.

2.1.2. Our approach

KerData's global approach consists in studying, designing, implementing and evaluating distributed algorithms and software architectures for scalable data storage and I/O management for efficient, large-scale data processing. We target two main execution infrastructures: cloud platforms and post-Petascale HPC supercomputers.

2.1.2.1. Platforms and Methodology

The highly experimental nature of our research validation methodology should be emphasized. To validate our proposed algorithms and architectures, we build software prototypes, then validate them at a large scale on real testbeds and experimental platforms.

We strongly rely on the Grid'5000 platform. Moreover, thanks to our projects and partnerships, we have access to reference software and physical infrastructures. In the cloud area, we use the Microsoft Azure and Amazon cloud platforms. In the post-Petascale HPC area, we are running our experiments on systems including some top-ranked supercomputers, such as Titan, Jaguar, Kraken or Blue Waters. This provides us with excellent opportunities to validate our results on advanced realistic platforms.

2.1.2.2. Collaboration strategy

Our collaboration portfolio includes international teams that are active in the areas of data management for clouds and HPC systems, both in Academia and Industry.

Our academic collaborating partners include Argonne National Lab, University of Illinois at Urbana-Champaign, Universidad Politécnica de Madrid, Barcelona Supercomputing Center, University Politehnica of Bucharest. In industry, we are mainly collaborating with Microsoft and IBM.

Moreover, the consortiums of our collaborative projects include application partners in the areas of Bio-Chemistry (e.g., IBCP Lyon in the MapReduce ANR project), Neurology and Genetics (e.g., the Parietal team at Inria, the NeuroSpin centre in Saclay within the A-Brain Microsoft Research-Inria project), and Climate Simulations (e.g., the Department of Earth and Atmospheric Sciences of the University of Michigan, within our collaboration inside JLESC [45]). This is an additional asset, which enables us to take into account application requirements in the early design phase of our solutions, and to validate those solutions with real applications... and real users!

3. Research Program

3.1. Research axis 1: Convergence of Extreme-Scale Computing and Big Data Infrastructures

The tools and cultures of High Performance Computing and Big Data Analytics have evolved in divergent ways. This is to the detriment of both. However, big computations still generate and are needed to analyze Big Data. As scientific research increasingly depends on both high-speed computing and data analytics, the potential interoperability and scaling convergence of these two eco-systems is crucial to the future. Our objective for the next years is premised on the idea that we must begin to systematically map out and account for the ways in which the major issues associated with Big Data intersect with, impinge upon, and potentially change the plans that are now being laid for achieving Exascale computing.

3.1.1. High-performance storage for concurrent Big Data applications

We argue that storage is a plausible pathway to convergence. In this context, we plan to focus on the needs of concurrent Big Data applications that require high-performance storage, as well as transaction support. Although blobs (binary large objects) are an increasingly popular storage model for such applications, state-of-the-art blob storage systems offer no transaction semantics. This demands users to coordinate data access carefully in order to avoid race conditions, inconsistent writes, overwrites and other problems that cause erratic behavior.

We argue there is a gap between existing storage solutions and application requirements, which limits the design of transaction-oriented applications. In this context, one idea on which we plan to focus our efforts is exploring how blob storage systems could provide built-in, multi-blob transactions, while retaining sequential consistency and high throughput under heavy access concurrency.

The early principles of this research direction have already raised interest from our partners at ANL (Rob Ross) and UPM (María Pérez) for potential collaborations. In this direction, the acceptance of our paper on the Týr transactional blob storage system as a Best Student Paper Award Finalist at the SC16 conference [25] is a very encouraging step.

3.1.2. Big Data analytics on Exascale HPC machines

Big Data analytics is another interesting direction that we plan to explore, building on top of these converged storage architectures. Specifically, we will examine the ways in which Exascale infrastructures can be leveraged not only by HPC-centric, but also by scientific, cloud-centric applications. Many of the current state-of-the-art Big Data processing approaches, including Hadoop and Spark [46] are optimized to run on commodity machines. This impacts the mechanisms used to deal with failures and the limited network bandwidth.

A blind adoption of these systems on extreme-scale platforms would result in high overheads. It would therefore prevent users from fully benefiting from the high performance infrastructure. The objective that we set here is to explore design and implementation options for new data analytics systems that can exploit the features of extreme-scale HPC machines: multi-core nodes, multiple memory and storage technologies including a large memory, NVRAM, SSDs, etc.

Collaboration. *This axis is addressed in close collaboration with María Pérez (UPM), Rob Ross (ANL), Toni Cortes (BSC), Bogdan Nicolae (formerly at IBM Research, now at Huawei Research).*

Relevant groups with similar interests are the following ones.

- *The group of Jack Dongarra, Innovative Computing Laboratory at University of Tennessee/Oak Ridge National Laboratory, working on joint tools Exascale Computing and Big Data.*
- *The group of Satoshi Matsuoka, Tokyo Institute of Technology, working on system software for Clouds and HPC.*
- *The group of Franck Cappello at Argonne National Laboratory/NCSA working on on-demand data analytics and storage for extreme-scale simulations and experiments.*

3.2. Research axis 2: Advanced data processing on Clouds

The recent evolutions in the area of Big Data processing have pointed out some limitations of the initial Map-Reduce model. It is well suited for batch data processing, but less suited for real-time processing of dynamic data streams. New types of data-intensive applications emerge, e.g., for enterprises who need to perform analysis on their stream data in ways that can give fast results (i.e., in real time) at scale (e.g., click-stream analysis and network-monitoring log analysis). Similarly, scientists require fast and accurate data processing techniques in order to analyze their experimental data correctly at scale (e.g., collectively analysis of large data sets distributed in multiple geographically distributed locations).

Our plan is to revisit current data management techniques to cope with the volatile requirements of data-intensive applications on large-scale dynamic clouds in a cost-efficient way.

3.2.1. Stream-oriented, Big Data processing on clouds

The state-of-the-art Hadoop Map-Reduce framework cannot deal with stream data applications, as it requires the data to be initially stored in a distributed file system in order to process them. To better cope with the above-mentioned requirements, several systems have been introduced for stream data processing such as Flink [41], Spark [46], Storm [47], and Google MillWheel [49]. These systems keep computation in memory to decrease latency, and preserve scalability by using data-partitioning or dividing the streams into a set of deterministic batch computations.

However, they are designed to work in dedicated environments and they do not consider the performance variability (i.e., network, I/O, etc.) caused by resource contention in the cloud. This variability may in turn cause high and unpredictable latency when output streams are transmitted to further analysis. Moreover, they overlook the dynamic nature of data streams and the volatility in their computation requirements. Finally, they still address failures in a best-effort manner.

Our objective is to investigate new approaches for reliable, stream Big Data processing on clouds. We will explore new mechanisms that expose resource heterogeneity (observed variability in resource utilization at runtime) when scheduling stream data applications. We will also investigate how to adapt to node failures automatically, and to adapt the failure handling techniques to the characteristics of the running application and to the root cause of failures.

3.2.2. Geographically distributed workflows on multi-site clouds

Many data processing jobs in data-intensive applications are modeled as workflows (i.e., as sets of tasks linked according to their data and computation dependencies) to facilitate the management and analysis of large volumes of data. With the fast growth of volumes of data to be handled at larger and larger scales, geographically distributed workflows are emerging as a natural data processing paradigm. This may bring several benefits: resilience to failures, distribution across partitions (e.g., moving computation close to data or vice versa), elastic scaling to support usage bursts, user proximity, etc.

In this context, sharing, disseminating and analyzing the data sets results in frequent large-scale data movements across widely distributed sites. Studies show that the inter-datacenter traffic is expected to triple in the following years. Our objective is to investigate approaches to data management enabling an efficient execution of such geographically distributed workflows running on multi-site clouds.

While in the past years we have addressed some data management issues in this area, mainly in support to efficient task scheduling of scientific workflows running on multisite clouds, we will now focus on an increasingly common scenario where workflows generate and process a huge number of small files, which is particularly challenging. As such workloads generate a deluge of small and independent I/O operations, efficient data and metadata handling is critical. We will explore specific means to better hide latency for data and metadata access in such scenarios, as a way to improve global performance.

Collaboration. *This axis is addressed in close collaboration with María Pérez (UPM), Kate Keahey (ANL) and Toni Cortes (BSC).*

Relevant groups with similar interests include the following ones.

- *The AMPLab, UC Berkeley, USA, working on scheduling stream data applications in heterogeneous clouds.*
- *The group of Ewa Deelman, USC Information Sciences Institute, working on resource management for workflows in Clouds.*
- *The XTRA group, Nanyang Technological University, Singapore, working on resource provisioning for workflows in the cloud.*

3.3. Research axis 3: I/O management, in situ visualization and analysis on HPC systems at extreme scales

Over the past few years, the increasing amounts of data produced by large-scale simulations have motivated a shift from traditional offline data analysis to in situ analysis and visualization. In situ processing started by coupling a parallel simulation with an analysis or visualization library, to avoid the cost of writing data on storage and reading it back. Going beyond this simple pairwise tight coupling, complex analysis workflows today are graphs with one or more data sources and several interconnected analysis components.

3.3.1. Toward a joint optimized architecture for in situ visualization and advanced processing

From Inria and ANL, four tools at least have emerged to address some challenges of coupling simulations with visualization packages or analysis workflows. Each of them focused on some particular aspect:

Damaris (Inria, [12], [4]) exploits dedicated cores to enable jitter-free I/O and in situ visualization;

Decaf (ANL, [36]) implements a coupling service for workflows;

FlowVR (Inria, [48]) connects workflow components for in situ processing;

Swift (ANL, [51]) focuses on implicitly parallel data flows and was optimized for Big Data processing.

Our plan is to explore how these tools could best leverage their respective strengths in a *joint optimized architecture for in situ visualization and advanced processing* in the HPC area. We published a preliminary study describing the lessons learned from using these tools in production environments with real applications [6]. Such a joint architecture will contribute to address the data volume and velocity challenges raised by data-intensive workflows, including complex data-intensive analytics phases. It may also impact, in a subsequent step, future data analysis pipelines for converged Big Data and HPC architectures.

Collaboration. *This axis is worked out in close collaboration with Rob Ross (ANL), Tom Peterka (ANL), Matthieu Dorier (ANL), Toni Cortes (BSC), Bruno Raffin (Inria). Some additional collaborations are in discussion with other members of JLESC, and with CEA and Total.*

Relevant groups with similar interests include the following ones.

- *The group of Manish Parashar at Rutgers University, USA (I/O management for HPC systems, in situ processing).*
- *The group of Scott Klasky at Oak Ridge National Lab, USA (I/O management for HPC systems, in situ processing).*
- *The CNRS IPSL laboratory (Sébastien Denvil, Pôle de modélisation du climat) in Paris, France (in situ data analytics).*

4. Application Domains

4.1. Application Domains

Our research work aims to improve large-scale, data-intensive applications running on clouds and extreme-scale HPC systems, with high requirements in terms of data storage and processing. Here are some classes of such applications.

Extreme-scale, data-intensive science simulations. A major research topic in the context of HPC simulations running on extreme-scale supercomputers is to explore how to record and visualize data during the simulation efficiently, without impacting the performance of the computation generating that data. In this area. We explore innovative approaches to I/O management and to in situ processing, in particular through our Damaris approach.

Map-Reduce-based data analytics. As Map-Reduce emerged as a dominant programming model for data analytics, we focus on several related challenges: how to enable fast failure recovery in shared Hadoop clusters; how to improve scheduling policies to favor resource allocation fairness; how to improve performance by detecting and mitigating stragglers.

Geographically-distributed cloud workflows. With fast-growing volumes of data to be handled at larger and larger scales, geographically distributed workflows are emerging as a natural data processing paradigm. They actually bring several benefits: resilience to failures, distribution across partitions, elastic scaling, user proximity etc. In this context, we investigate approaches to data management enabling an efficient execution of such geographically distributed workflows running on multi-site clouds. In projects like *ANR OverFlow* and *Z-CloudFlow* we explore means to better hide latency for data and metadata access and optimize transfers as a way of improving the global performance.

Stream data processing. The evolutions in the area of Big Data processing, the development of cloud computing and the success of the Map-Reduce model have fostered new types of data-intensive applications, in which obtaining fast and timely results is mandatory. Enterprises need to perform analysis on their stream data that can give fast results (i.e., in real time) at scale (e.g., click-stream analysis and network-monitoring log analysis). Similarly, scientists require fast and accurate data processing techniques in order to analyze their experimental data correctly at scale (e.g., analysis of data produced by massive-scale simulations and sensor deployments).

Besides processing, we are also focusing on efficient stream data storage. Unlike traditional storage, the main challenge of storing stream data is the large number of small items (arriving at rates easily reaching tens of millions per second). We explore the plausible paths towards a dedicated storage solution. We aim to provide on the one hand traditional storage functionality, and on the other hand stream-like performance (i.e., low-latency I/O access to items and ranges of items).

The team's projects and collaborations explicitly target concrete use cases belonging to the above application classes, in the following areas.

Smart Cities and Territories. In the framework on the *BigStorage project* where the KerData team is a major partner, we are focusing on several stream data applications in the context of Smart cities. The goal is to optimize current state-of-the-art processing engines to provide real-time analyzing of data collected from small sensors and devices. This will enable to make smart decisions in fields like healthcare, traffic management, water quality, air pollution and many more.

Climate and meteorology. An example is the atmospheric simulation code CM1 (Cloud Model 1), one of the target applications of the Blue Waters machine. We already used this code in collaborative research within *Data@Exascale* Associate Team, in the framework of the *Joint Laboratory for Extreme-Scale Computing* (JLESC), co-supported by Inria, UIUC, ANL, BSC, JSC and RIKEN/AICS.

Brain imaging. In the *A-Brain* MSR-Inria project (now completed), we applied Map-Reduce-based data analytics to neuro-imaging genetics.

Molecular biology. In the framework of the *MapReduce ANR project* led by KerData (now completed), we have focused on the *FastA* bioinformatics application used for massive protein sequence similarity searching. In the context of the *OverFlow ANR project* we are pursuing this analysis in collaboration with the Institut Français de Bioinformatique (IFB).@ We aim at using these results for drug design in an industrial context (i.e. the identification of new druggable protein targets and thereby the generation of new drug candidates).

5. Highlights of the Year

5.1. Highlights of the Year

5.1.1. Awards

SC16: Best Student Paper Finalist. The paper entitled *Týr: Blob Storage Meets Built-In Transactions* presented by Pierre Matri at the Supercomputing (SC16) Conference was one of the 7 papers selected for the Best Student Paper award.

This work was carried out in the context of the BigStorage project, under the supervision of Alexandru Costan, Gabriel Antoniu, María Pérez , and Jesús Montes.

There were 442 submissions, and 81 accepted papers.

ACM Graduate Student Research Competition SC16. Nathanaël Cheriere received the third prize in the SC16 ACM Student Research Competition for his work on optimizing the algorithms for the MPI collective *Scatter* and *AllGather* routines on the Dragonfly topology [1].

This work was carried out at the Argonne National Laboratory in the context of the JLESC, under the supervision of Matthieu Dorier, Rob Ross, Shadi Ibrahim, and Gabriel Antoniu.

As many as 62 posters were submitted for the Student Research Competition, out of which 14 have been selected in the Graduate category. After the presentation of their posters, 4 students have been invited to make a presentation of their work in front of a jury.

5.1.2. 9 papers in international journals

This year the team published 9 papers in high-quality journals including *ACM Transactions on Parallel Computing*, *IEEE Transactions on Parallel and Distributed Systems*, *Future Generation Computer Systems*, *Concurrency and Computation: Practice and Experience* and *IEEE Transactions on Cloud Computing*.
BEST PAPERS AWARDS :
[25] **IEEE ACM SC16 - The International Conference for High Performance Computing, Networking, Storage and Analysis 2016**. P. MATRI, A. COSTAN, G. ANTONIU, J. MONTES, M. S. PÉREZ.

6. New Software and Platforms

6.1. Týr

Title: Týr: Blob Storage Meets Built-In Transactions.

Keywords: Big Data; Transactions; Tyr; BlobSeer.

Scientific Description: Týr [25] is the first blob storage system to provide built-in, multi-blob transactions, while retaining sequential consistency and high throughput under heavy access concurrency.

Functional Description: Týr offers fine-grained random write access to data and in-place atomic operations. Large-scale experiments on Microsoft Azure with a production application from CERN LHC show Týr throughput outperforms state-of-the-art solutions by more than 75%. Týr leverages the approaches developed within BlobSeer, the reference data management system for large distributed blobs, developed over the past years in KerData.

Contact data:

Participants: Pierre Matri, Alexandru Costan and Gabriel Antoniu.

Partners: INSA Rennes, Universidad Politécnica de Madrid.

Contact: Gabriel Antoniu.

URL: http://tyr.io/.

6.2. Damaris

Title: Damaris: I/O and data management for large-scale, MPI-based HPC simulations.

Keywords: I/O; HPC; Data management; Visualization; Big Data; Exascale.

Scientific Description: Damaris is a middleware for multicore SMP nodes enabling them to efficiently handle data transfers for storage and visualization. The key idea is to dedicate one or a few cores of each SMP node to the application I/O. It is developed within the framework of a collaboration between KerData and the JLESC. The current version enables efficient asynchronous I/O, hiding all I/O-related overheads such as data compression and post-processing, as well as direct (in situ) interactive visualization of the generated data.

Damaris has been preliminarily evaluated at NCSA (Urbana-Champaign) with the CM1 tornado simulation code. CM1 is one of the target applications of the Blue Waters supercomputer in production at NCSA/UIUC (USA), in the framework of the JLESC. Damaris now has external users, including (to our knowledge) visualization specialists from NCSA, Big Data experts from the HDF group, and researchers from the France/Brazil Associated Research Team on Parallel Computing (joint team between Inria/LIG Grenoble and the UFRGS in Brazil). Damaris has been successfully integrated into four large-scale simulations (CM1, OLAM, Nek5000, CROCO). Works are in progress to evaluate it in the context of several other simulation codes including HACC (cosmology) and GTC (fusion).

Damaris is the object of a *Technical Development Action* (ADT) supported by Inria.

Functional Description: Damaris targets large-scale HPC simulations: in situ data analysis by some dedicated cores of the simulation platform; asynchronous and fast data transfer from HPC simulations to Damaris; semantic-aware dataset processing through Damaris plug-ins.

Contact data:

Participants: Matthieu Dorier (ANL), Lokman Rahmani, Gabriel Antoniu, Orçun Yildiz, Hadi Salimi and Luc Bougé.

Partners: ENS Rennes, Argonne National Laboratory.

Contact: Gabriel Antoniu.

URL: http://damaris.gforge.inria.fr/.

6.3. Other software

6.3.1. JetStream

Title: JetStream: Enabling High-Performance Event Streaming across Cloud Data-Centers.

Keywords: Big Data, streaming, data transfer, multisite cloud.

Scientific Description. JetStream is a middleware solution for batch-based, high-performance streaming across cloud data centers. JetStream implements a set of context-aware strategies to optimize batch-based streaming, being able to self-adapt to changing conditions.

Functional Description. The system provides multi-route streaming across cloud data centers for aggregating bandwidth by leveraging the network parallelism. It enables easy deployment across .Net frameworks and seamless binding with event processing engines such as StreamInsight. JetStream is currently used at Microsoft Research ATLE Munich for the management of the Azure cloud infrastructure.

Participants: Ovidiu-Cristian Marcu, Alexandru Costan, Gabriel Antoniu.

Contact: Alexandru Costan.

6.3.2. Omnisc'IO

Title: Omnisc'IO: a Grammar-Based Approach to Spatial and Temporal I/O Patterns Prediction.

Keywords: HPC, Input-Output, Prediction, Grammar.

Scientific Description. Omnisc'IO is a library that aims to be integrated into I/O middleware.

Functional Description. It traces I/O operations, models the stream of such operations using grammar-inference techniques, and predicts when new I/O operations will be performed, as well as where and how much data will be written.

Participants: Matthieu Dorier (ANL), Gabriel Antoniu, Shadi Ibrahim.

Contact: Gabriel Antoniu.

6.3.3. OverFlow

Title: OverFlow: Workflow Data Management as a Service for Multi-Site Applications.

Keywords: Small data; workflow; multi-site cloud.

Scientific Description. OverFlow is a uniform data management system for scientific workflows running across geographically distributed sites, aiming to reap economic benefits from this geo-diversity. The software is environment-aware, as it monitors and models the global cloud infrastructure, offering high and predictable performance for transfer cost and time, within and across sites.

Functional Description. OverFlow proposes a set of pluggable services, grouped in a data-scientist cloud kit. They provide the applications with the possibility to monitor the underlying infrastructure, to exploit smart data compression, deduplication and geo-replication, to evaluate data management costs, to set a tradeoff between money and time, and optimize the transfer strategy accordingly. Currently, OverFlow is used for data transfers by the Microsoft Research ATLE Munich team as well as for synthetic benchmarks at the Politehnica University of Bucharest.

Participants: Paul Le Noac'h, Ovidiu-Cristian Marcu, Alexandru Costan and Gabriel Antoniu.

Contact: Alexandru Costan.

6.3.4. iHadoop

Title: iHadoop: A Hadoop Simulator Developed In Java on Top of SimGrid.

Keywords: Simulation, Map-Reduce, Hadoop, SimGrid.

Scientific Description. iHadoop is a Hadoop simulator developed in Java on top of SimGrid. It simulates the behavior of Hadoop and therefore accurately predicts the performance of Hadoop in normal scenarios and under failures. iHadoop is extended to (1) simulate the execution and predict the performance of multiple Map-Reduce applications; (2) simulate the execution of Map-Reduce applications under various data distributions and data skew models.

Functional Description. iHadoop is an internal software prototype, which was initially developed to validate our idea regarding the behavior of Hadoop under failures. iHadoop has preliminarily evaluated within our group and it has shown very high accuracy to predict the execution time of a Map-Reduce applications. We intend to integrate iHadoop within the SimGrid distribution and make it available to the SimGrid community.

Participants: Shadi Ibrahim and Tien-Dat Phan.

Contact: Shadi Ibrahim.

7. New Results

7.1. Convergence of HPC and Big Data

7.1.1. Transactional storage

Participants: Pierre Matri, Alexandru Costan, Gabriel Antoniu.

Concurrent Big Data applications often require high-performance storage, as well as ACID (Atomicity, Consistency, Isolation, Durability) transaction support. Although blobs (binary large objects) are an increasingly popular model for addressing the storage needs of such applications, state-of-the-art blob storage systems typically offer no transaction semantics. This demands users to coordinate access to data carefully in order to avoid race conditions, inconsistent writes, overwrites and other problems that cause erratic behavior. We argue there is a gap between existing storage solutions and application requirements, which limits the design of transaction-oriented applications.

Týr is the first blob storage system to provide built-in, multi-blob transactions, while retaining sequential consistency and high throughput under heavy access concurrency. Týr offers fine-grained random write access to data and in-place atomic operations.

Large-scale experiments on Microsoft Azure with a production application from CERN LHC show Týr throughput outperforms state-of-the-art solutions by more than 75 %.

Collaboration. *This work was done in collaboration with María Pérez, UPM, Spain.*

7.1.2. Big Data on HPC

Participants: Orçun Yildiz, Shadi Ibrahim, Gabriel Antoniu.

Over the last decade, Map-Reduce has stood as the most powerful Big Data processing model. Map-Reduce model is now used by many companies and research labs to facilitate large-scale data analysis. With the growing needs of users and size of data, commodity-based infrastructure (most commonly used as of now) will strain under the heavy weight of Big Data. On the other hand, HPC systems offer a rich set of opportunities for Big Data processing.

As first steps towards Big Data processing on HPC systems, several research efforts have been devoted to understand Map-Reduce performance on these systems. Yet, the impact of the specific features of HPC environments have not been fully investigated, yet.

We conducted an experimental campaign to provide a clearer understanding of Map-Reduce performance on HPC systems. We use Spark, a widely adopted Map-Reduce framework, and representative Big Data workloads on Grid'5000 testbed to evaluate how the latency, contention and file system's configuration can influence the application performance.

7.1.3. Energy vs. performance trade-offs

Participants: Mohammed-Yacine Taleb, Shadi Ibrahim, Gabriel Antoniu.

Most large popular web applications, like Facebook and Twitter, have been relying on large amounts of in-memory storage to cache data and provide a low response time. As the memory capacity of clusters and clouds increases, it becomes possible to keep most of the data in the main memory.

This motivates the introduction of in-memory storage systems. While prior work has focused on how to exploit the low latency of in-memory access at scale, there is still little knowledge regarding the energy efficiency of in-memory storage systems. This is unfortunate, as it is known that main memory is a major energy bottleneck in many computing systems. For instance, DRAM consumes up to 40 % of a server's power.

By means of experimental evaluation, we have studied the performance and energy-efficiency of RAMCloud — a well-known in-memory storage system. We demonstrated that although RAMCloud is scalable for read-only applications, it exhibits non-proportional power consumption. We also found that the current replication scheme implemented in RAMCloud limits the performance and results in high energy consumption. Surprisingly enough, we also showed that replication can even play a negative role in crash-recovery.

Collaboration. *This work was carried out in collaboration with Toni Cortes (BSC, Spain).*

7.2. Efficient I/O and communication for Extreme-scale HPC systems

7.2.1. Adaptive performance-constrained in situ visualisation

Participant: Lokman Rahmani.

While many parallel visualization tools now provide in situ visualization capabilities, the trend has been to feed such tools with large amounts of unprocessed output data and let them render everything at the highest possible resolution. This leads to an increased run time of simulations that still have to complete within a fixed-length job allocation.

We have been working on tackling the challenge of enabling in situ visualization under performance constraints. Our approach shuffles data across processes according to their contents and filters out part of them. Thereby, the visualization pipeline is only fed with a reorganized subset of the data produced by the simulation.

Our framework, as presented in [22], leverages fast, generic evaluation procedures to score blocks of data, using information theory, statistics, and linear algebra. It monitors its own performance and dynamically adapts to achieve appropriate visual fidelity within predefined performance constraints. Experiments on the Blue Waters supercomputer with the CM1 simulation show that our approach enables a 5-time speedup with respect to the initial visualization pipeline, and is able to meet performance constraints.

Collaboration. *This was was carried out with the collaboration of Matthieu Dorier, ANL, USA.*

7.2.2. Dragonfly

Participants: Nathanaël Cheriere, Shadi Ibrahim, Gabriel Antoniu.

High-radix direct network topologies such as Dragonfly have been proposed for Petascale and Exascale supercomputers. It has been shown that they ensure fast interconnections and reduce the cost of the network compared to traditional network topologies. However, current algorithms for communication do not consider the topology and thus waste numerous opportunities of optimization for performance.

In our studies, we exploit the strength of the Dragonfly with topology-aware algorithms for AllGather and Scatter operations. We analyze existing algorithms, then propose derived algorithms, that we evaluate using CODES, an event-driven simulator.

As expected, making AllGather algorithms topology-aware does improve the performance and reduces the link utilization. However, simulations of various Scatter algorithms show surprising results, and point out the important role played by hardware for the efficiency of the algorithms. In particular, the knowledge of the number and size of input-output buffers in routers can be exploited to accelerate the Scatter operation by a factor up to 2 times.

Collaboration. *This work was done in collaboration with Matthieu Dorier and Rob Ross, ANL, USA.*

7.2.3. Interference between HPC jobs

Participants: Orçun Yildiz, Shadi Ibrahim, Gabriel Antoniu.

As we move toward the Exascale era, performance variability in HPC systems remains a challenge. I/O interference, a major cause of this variability, is becoming more important every day with the growing number of concurrent applications that share larger machines. Earlier research efforts on mitigating I/O interference focus on a single potential cause of interference (e.g., the network). Yet the root causes of I/O interference can be diverse.

In [27], we conducted an extensive experimental campaign to explore the various root causes of I/O interference in HPC storage systems. We used micro-benchmarks on the Grid'5000 testbed to evaluate how I/O interference is influenced by the applications' access pattern, the network components, the file system's configuration, and the backend storage devices.

Our studies revealed that in many situations interference is a result of a bad flow control in the I/O path, rather than being caused by some single bottleneck in one of its components. We further show that interference-free behavior is not necessarily a sign of optimal performance. To the best of our knowledge, our work provides the first deep insight into the role of each of the potential root causes of interference and their interplay. Our findings can help developers and platform owners improve I/O performance and motivate further research addressing the problem across all components of the I/O stack.

Collaboration. *This work was done in collaboration with Matthieu Dorier and Rob Ross, ANL, USA.*

7.3. Workflow on clouds

7.3.1. Managing hot metadata for scientific workflows on multisite clouds

Participants: Luis Eduardo Pineda Morales, Alexandru Costan, Gabriel Antoniu.

Large-scale scientific applications are often expressed as workflows that help defining data dependencies between their different components. Such workflows may incur huge storage and computation requirements, so that they need to be processed in multiple (cloud-federated) datacenters. A major challenge in such multisite clouds is the long latency of the network links between datacenters, that limits the performance of multisite applications. Moreover, it has been shown that poor metadata handling can further impact the efficiency of computing systems. Many efforts have been done to improve metadata management; however, most of them concern only single-site, HPC systems to date.

In [26], we assert that some workflow metadata are more frequently accessed than other, and thus should be handled with higher priority during the workflow's execution. We call them *hot metadata*. We present a hybrid decentralized/distributed model for handling hot metadata in *multisite* architectures. We couple our model with a scientific workflow management system (SWfMS) to validate and tune its applicability to various real-life scientific scenarios. We show that efficient management of hot metadata improves the performance of SWfMS, reducing the workflow execution time up to 50 % for highly parallel jobs by enabling timely data provisioning and avoiding unnecessary *cold* metadata operations.

7.3.2. *Probabilistic optimizations for resource provisioning of cloud workflows*

Participants: Chi Zhou, Shadi Ibrahim.

In many data-intensive applications, data management routines can be represented as workflows, where tasks are organized according to data and computation dependencies. Recently, the optimal provisioning of resources (e.g., VMs) for workflows running in the cloud has attracted a lot of attention. Most resource provisioning solutions overlook the important factor of cloud dynamics, e.g., the fluctuation of I/O, network performance, and system failures. In our experiments on the Amazon EC2 cloud, these issues significantly impact resource allocation quality. Therefore, we study how cloud dynamics should be incorporated into the resource provisioning process.

Our approach models cloud dynamics as time-dependent random variables (e.g., a probability distribution of workflow execution times) and performs probabilistic optimizations for resource provisioning problems using those random variables as optimization input. This solution yields more effective resource provisioning for cloud workflows. However, it involves heavy computation effort due to the complex structures of workflows and the large number of probability calculations.

To overcome this problem, we develop a three-stage pruning process, which simplifies workflow structure and reduces probability evaluation overhead. We have also implemented our techniques in a runtime library, which allows users to integrate our techniques into their existing resource provisioning methods. Experiments on two common resource provisioning problems show that probabilistic solutions can improve the performance by 51 % —70 % compared with state-of-the-art, static solutions.

Collaboration. *This work was done in collaboration with Bingsheng He NUS, Singapore.*

7.3.3. *A taxonomy and survey of scientific computing in the cloud*

Participants: Chi Zhou, Shadi Ibrahim.

Cloud computing has evolved as a popular computing infrastructure for many applications. With (big) data acquiring a crucial role in eScience, efforts have been made recently to develop and deploy scientific applications efficiently on the unprecedentedly scalable cloud infrastructures.

In [29], we review recent efforts in developing and deploying scientific computing applications in the cloud. In particular, we introduce a taxonomy specifically designed for scientific computing in the cloud, and further review the taxonomy with four major kinds of science applications, including life sciences, physics sciences, social and humanities sciences, and climate and earth sciences.

Due to the large data size in most scientific applications, the performance of I/O operations can greatly affect the overall performance of the applications. As a consequence, the dynamic I/O performance of the cloud has made resource provisioning an important and complex problem for scientific applications in the cloud.

We present our efforts on improving the resource provisioning efficiency and effectiveness of scientific applications in the cloud. Finally, we present the open problems for developing the next-generation eScience applications and systems in the cloud and give our conclusions.

Collaboration. *This work was done in collaboration with Bingsheng He NUS, Singapore.*

7.4. Fault tolerant data processing

7.4.1. Fast recovery

Participants: Orçun Yildiz, Shadi Ibrahim, Gabriel Antoniu.

Hadoop has emerged as a prominent tool for Big Data processing in large-scale clouds. Failures are inevitable in large-scale systems, especially in shared environments. Consequently, Hadoop was designed with hardware failures in mind. In particular, Hadoop handles machine failures by re-executing all the tasks of the failed machine. Unfortunately, the efforts to handle failures are entirely entrusted to the core of Hadoop and hidden from Hadoop schedulers. This may prevent Hadoop schedulers from meeting their objectives (e.g., fairness, job priority, performance) and can significantly impact the performance of the applications.

In our previous work, we addressed this issue through the design and implementation of a new scheduling strategy called Chronos. Chronos is conductive to improving the performance of Map-Reduce applications by enabling an early action upon failure detection. Chronos tries to launch recovery tasks immediately by preempting tasks belonging to low priority jobs, thus avoiding to wait until slots are freed.

In [20], we further investigated the potential benefit of launching local recovery tasks by implementing and evaluating Chronos*. To this end, we slightly changed the smart slot allocation strategy of Chronos into aggressive slot allocation strategy. With Chronos, recovery tasks with higher priority would preempt the selected tasks with less priority. With Chronos*, we also allow recovery tasks to preempt the selected tasks with the same priority (e.g., recovery tasks belonging to the same job with selected tasks). The experimental results indicate that Chronos* results in 100 % locality execution for recovery tasks thanks to its aggressive slot allocation strategy. Moreover, Chronos* improves the completion time of the jobs by up to 17 %.

7.4.2. Dynamic replica placement

Participants: Pierre Matri, Alexandru Costan, Gabriel Antoniu.

Large-scale applications are ever-increasingly geo-distributed. Maintaining the highest possible *data locality* is crucial to ensure high performance of such applications. Dynamic replication addresses this problem by dynamically creating replicas of frequently accessed data close to the clients. This data is often stored in decentralized storage systems such as Dynamo or Voldemort, which offer support for *mutable data*.

However, existing approaches to dynamic replication for such mutable data remain centralized, thus incompatible with these systems. We introduce a write-enabled dynamic replication scheme that leverages the decentralized architecture of such storage systems. We propose an algorithm enabling clients to locate tentatively the closest data replica without prior request to any metadata node. Large-scale experiments show a read latency decrease of up to 42% compared to other state-of-the-art, caching-based solutions.

Collaboration. *This work was done in collaboration with María Pérez, UPM, Spain.*

7.5. Advanced data management on clouds

7.5.1. Benchmarking Spark and Flink

Participants: Ovidiu-Cristian Marcu, Alexandru Costan, Gabriel Antoniu.

Spark and Flink are two Apache-hosted data analytics frameworks that represent the state of the art in modern in-memory Map-Reduce processing. They facilitate the development of multi-step data pipelines using directly acyclic graph (DAG) patterns. In the framework of our BigStorage project, we performed a comparative study [23] which evaluates the performance of Spark versus Flink. The objective is to identify and explain the impact of the different architectural choices and the parameter configurations on the perceived end-to-end performance.

Based on empirical evidences, the study points out that in Big Data processing there is not a single framework for all data types, sizes and job patterns and emphasize a set of design choices that play an important role in the behaviour of a Big Data framework: memory management, pipelined execution, optimizations and parameter configuration easiness. What raises our attention is that a streaming engine (i.e., Flink) delivers in many benchmarks better performance than a batch-based engine (i.e., Spark), showing that a more general Big Data architecture (treating batches as finite sets of streamed data) is plausible and may subsume both streaming and batching use cases.

Collaboration. *This work was done in collaboration with María Pérez, UPM, Spain.*

7.5.2. Geo-distributed graph processing

Participants: Chi Zhou, Shadi Ibrahim.

Graph processing is an emerging model adopted by a wide range of applications to easily parallelize the computations over graph data. Partitioning graph processing workloads to multiple machines is an important task for reducing the communication cost and improving the performance of graph processing jobs. Recently, many real-world applications store their data on multiple geographically distributed datacenters (DCs) to ensure flexible and low-latency services. Due to the limited Wide Area Network (WAN) bandwidths and the network heterogeneity of the geo-distributed DCs, existing graph partitioning methods need to be redesigned to improve the performance of graph processing jobs in geo-distributed DCs.

To address the above challenges, we propose a heterogeneity-aware graph partitioning method named G-Cut, which aims at minimizing the runtime of graph processing jobs in geo-distributed DCs while satisfying the WAN usage budget. G-Cut is a two-stage graph partitioning method. In the traffic-aware graph partitioning stage, we adopt the one-pass edge assignment to place edges into different partitions while minimizing the inter-DC data traffic size. In the network-aware partition refinement stage, we map the partitions obtained in the first stage onto different DCs in order to minimize the inter-DC data transfer time. We evaluate the effectiveness and efficiency of G-Cut using real-world graphs and the evaluation results show that G-Cut can achieve both lower WAN usage and shorter inter-DC data transfer time compared to state-of-the-art graph partitioning methods.

Collaboration. *This work was done in collaboration with Bingsheng He NUS, Singapore.*

7.5.3. Fairness and scheduling

Participants: Orçun Yildiz, Shadi Ibrahim, Gabriel Antoniu.

Recently, Map-Reduce and its open-source implementation Hadoop have emerged as prevalent tools for big data analysis in the cloud. Fair resource allocation in-between jobs and users is an important issue, especially in multi-tenant environments such as clouds. Several scheduling policies have been developed to preserve fairness in multi-tenant Hadoop clusters. At the core of these schedulers, simple (non-) preemptive approaches are employed to free resources for tasks belonging to jobs with less share. For example, Hadoop Fair Scheduler is equipped with two approaches: wait and kill. While wait may introduce a serious violation in fairness, kill may result in a huge waste of resources. Yet, recently some work have introduced preemption approach in shared Hadoop clusters.

To this end, we closely examine three approaches including wait, kill and preemption when Hadoop Fair Scheduler is employed for ensuring fair execution between multiple concurrent jobs. We perform extensive experiments to assess the impact of these approaches on performance and resource utilization while ensuring fairness. Our experimental results bring out the differences between these approaches and illustrate that these approaches are only sub-optimal for different workloads and cluster configurations: the efficiency of achieving fairness and the overall performance varies with the workload composition, resource availability and the cost of the adopted preemption technique.

7.5.4. Stragglers in Map-Reduce

Participants: Tien-Dat Phan, Shadi Ibrahim.

Big Data systems (e.g., Map-Reduce, Hadoop, Spark) rely increasingly on speculative execution to mask slow tasks also known as stragglers because a job's execution time is dominated by the slowest task instance. Big Data systems typically identify stragglers and speculatively run copies of those tasks with the expectation a copy may complete faster to shorten job execution times.

There is a rich body of recent results on straggler mitigation in Map-Reduce. However, the majority of these do not consider the problem of accurately detecting stragglers. Instead, they adopt a particular straggler detection approach and then study its effectiveness in terms of performance, e.g., reduction in job completion time, or its efficiency, e.g., extra resource usage.

In this work, we consider a complete framework for straggler detection and mitigation. We start with a set of metrics that can be used to characterizes and detect stragglers such as Precision, Recall, Detection Latency, Undetected Time and Fake Positive. We then develop an architectural model by which these metrics can be linked to measures of performance including execution time and system energy overheads.

We further conduct a series of experiments to demonstrate which metrics and approaches are more effective in detecting stragglers and are also predictive of effectiveness in terms of performance and energy efficiency. For example, our results indicate that the default Hadoop straggler detector could be made more effective. In certain cases, precision is low and only 65 % of those detected are actual stragglers and recall, i.e., the proportion of stragglers which are actually detected, is also relatively low at 56 %. For the same case, the hierarchical approach (i.e., a green-driven detector based on the default one) achieves a precision of 98 % and a recall of 33 %.

Further, these increases in precision can be used to achieve lower execution time and energy consumption, and thus higher performance and energy efficiency. Compared to the default Hadoop mechanism, energy consumption is reduced by almost 30 %. These results demonstrate how our framework can offer useful insights and be applied in practical settings to characterize and design new straggler detection mechanisms for Map-Reduce systems.

Collaboration. *This work was carried out in collaboration with Guillaume Aupy and Padma Raghavan whilst they were affiliated with Vanderbilt University, USA.*

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Contracts with Industry

Microsoft: Z-CloudFlow (2013–2016). In the framework of the Joint Inria-Microsoft Research Center, this project is a follow-up to the A-Brain project. The goal of this new project is to propose a framework for the efficient processing of scientific workflows in clouds. This approach will leverage the cloud infrastructure capabilities for handling and processing large data volumes.

In order to support data-intensive workflows, the cloud-based solution will: adapt the workflows to the cloud environment and exploit its capabilities; optimize data transfers to provide reasonable times; manage data and tasks so that they can be efficiently placed and accessed during execution.

The validation will be performed using real-life applications, first on the Grid5000 platform, then on the Azure cloud environment, access being granted by Microsoft through a *Azure for Research Award* received by G. Antoniu. The project also provides funding for the PhD thesis of Luis Pineda-Morales, started in 2014.

Collaboration. *The project is being conducted in collaboration with the Zenith team from Montpellier, led by Patrick Valduriez.*

Huawei: HIRP Low-Latency Storage for Stream Data (2016–2017). The goal of this project is to explore the plausible paths towards a dedicated storage solution for low-latency stream storage. Such a solution should provide on the one hand traditional storage functionality and on the other hand stream-like performance (i.e., low-latency I/O access to items and ranges of items).

We plan to investigate the main requirements and challenges, evaluate the different design choices (e.g., a standalone component vs. an extension of an existing Big Data solution like HDFS) and then propose an architectural overview.

9. Partnerships and Cooperations

9.1. National Initiatives

9.1.1. ANR

9.1.1.1. OverFlow (2015–2019)

- Project Acronym: OverFlow.
- Project Title: Workflow Data Management as a Service for Multisite Applications.
- Coordinator: Alexandru Costan.
- Duration: Octobre 2015–October 2019.
- Other Partners: None (Young Researcher Project).
- External collaborators: Kate Keahey (University of Chicago and Argonne National Laboratory), Bogdan Nicolae (Huawei Research) and Christophe Blanchet (Institut Français de Bioinformatique).
- Abstract: This JCJC project led by Alexandru Costan investigates approaches to data management enabling an efficient execution of geographically distributed workflows running on multi-site clouds. Ultimately, OverFlow will propose a new, pioneering paradigm: Workflow Data Management as a Service — a general and easy-to-use, cloud-provided service that bridges for the first time the gap between single- and multi-site workflow data management. It aims to reap economic benefits from the geo-diversity while accelerating the scientific discovery through a democratization of access to globally distributed data.

9.1.2. Other National Projects

9.1.2.1. DISCOVERY (2015–2019)

- Project Acronym: DISCOVERY.
- Project Title: DIStributed and COoperative framework to manage Virtual EnviRonments autonomicallY.
- Coordinator: Adrien Lèbre.
- Duration: 2015–2019.
- Partners: Inria Project-Teams including ASAP, ASCOLA, Avalon, Myriads, and KerData.
- Abstract: An Inria Project Lab, led by Adrien Lèbre (ASCOLA), that aims at exploring a new way of operating Utility Computing (UC) resources by leveraging any facilities available through the Internet. The goal is to deliver widely distributed platforms that can better match the geographical dispersal of users, as well as the unending demand.

  Within DISCOVERY, S. Ibrahim (KerData Inria Team) is working with Gilles Fedak (Avalon Inria Project-Team) to address the VM images management challenge.

9.1.2.2. ADT Damaris

- Project Acronym: ADT Damaris

- Project Title: Technology development action for te Damaris environment.
- Coordinator: Alexandru Costan.
- Duration: 2016–2018.
- Abstract: This action aims to support the development of the Damaris software. Inria's *Technological Development Office* (D2T, *Direction du Développement Technologique*) provided 2 years of funding support for a senior engineer.

  Hadi Salimi is funded through this project to document, test and extend the Damaris software and make it a safely distributable product.

9.1.2.3. Grid'5000.

We are members of Grid'5000 community and run experiments on the Grid'5000 platform on a daily basis.

9.2. European Initiatives

9.2.1. FP7 and H2020 Projects

9.2.1.1. BigStorage

- Title: BigStorage: Storage-based Convergence between HPC and Cloud to handle Big Data.
- Programme: H2020.
- Duration: January 2015–December 2018.
- Coordinator: Universidad Politécnica de Madrid (UPM).
- Partners:
  - Barcelona Supercomputing Center — Centro Nacional de Supercomputacion (Spain)
  - CA Technologies Development Spain (Spain)
  - CEA — Commissariat à l'énergie atomique et aux énergies alternatives (France)
  - Deutsches Klimarechenzentrum (Germany)
  - Foundation for Research and Technology Hellas (Greece)
  - Fujitsu Technology Solutions (Germany)
  - Johannes Gutenberg Universitaet Mainz (Germany)
  - Universidad Politecnica de Madrid (Spain)
  - Seagate Systems UK (United Kingdom)
- Inria contact: G. Antoniu and Adrien Lèbre.
- URL: http://www.bigstorage-project.eu/.
- Description: BigStorage is a European Training Network (ETN) whose main goal is to train future *data scientists*. It aims at enabling them and us to apply holistic and interdisciplinary approaches to take advantage of a data-overwhelmed world. This world requires *HPC* and *Cloud* infrastructures with a redefinition of *storage* architectures underpinning them — focusing on meeting highly ambitious performance and *energy* usage objectives. The KerData team will be hosting 2 *Early Stage Researchers* in this framework.

9.3. International Initiatives

9.3.1. Inria International Labs

9.3.1.1. JLESC: Joint Laboratory on Extreme-Scale Computing

The Joint Laboratory on Extreme-Scale Computing is jointly run by Inria, UIUC, ANL, BSC, JSC and RIKEN/AICS. It has ben created in 2014 as a follow-up of the Inria-UIUC JLPC, the *Joint Laboratory for Petascale Computing*.

The KerData team is collaborating with teams from ANL and UIUC within this lab since 2009 on several topics in the areas of I/O, storage and in situ processing and cloud computing. This collaboration has been initially formalized as the *Data@Exascale* Associate Team with ANL and UIUC (2013–2015) followed by *Data@Exascale 2* Associate Team with ANL (2016–2018).

Since 2015, Gabriel Antoniu serves as a topic leader for Inria for the *I/O, Storage and In Situ Processing* topic.

9.3.1.1.1. Associate Team involved in the International Lab: Data@Exascale 2

Project Acronym: Data@Exascale 2.

Project Title: Convergent Data Storage and Processing Approaches for Exascale Computing and Big Data Analytics.

International Partner:

- Argonne National Laboratory (United States) — Mathematics and Computer Science Division (MCS) — Rob Ross.

Start year: 2013.

URL: http://www.irisa.fr/kerdata/data-at-exascale/.

Description: In the past few years, countries including United States, the European Union, Japan and China have set up aggressive plans to get closer to what appears to be the next goal in terms of high-performance computing (HPC): Exaflop computing, a target which is now considered reachable by the next-generation supercomputers in 2020-2023. While these government-led initiatives have naturally focused on the big challenges of Exascale for the development of new hardware and software architectures, the quite recent emergence of the Big Data phenomenon introduces what could be called a tectonic shift that is impacting the entire research landscape for Exascale computing. As data generation capabilities in most science domains are now growing substantially faster than computational capabilities, causing these domains to become data-intensive, new challenges appeared in terms of volumes and velocity for data to be stored, processed and analyzed on the future Exascale machines.

To face the challenges generated by the exponential data growth (a general phenomenon in many fields), a certain progress has already been made in the recent years in the rapidly-developing, industry-led field of cloud-based Big Data analytics, where advanced tools emerged, relying on machine-learning techniques and predictive analytics.

Unfortunately, these advances cannot be immediately applied to Exascale computing: the tools and cultures of the two worlds, HPC (High-Performance Computing) and BDA (Big Data Analytics) have developed in a divergent fashion (in terms of major focus and technical approaches), to the detriment of both. The two worlds share however multiple similar challenges and unification now appears as essential in order to address the future challenges of major application domains that can benefit from both.

The scientific program we propose for the Data@Exascale 2 Associate Team is defined from this new, highly-strategic perspective and builds on the idea that the design of innovative approaches to data I/O, storage and processing allowing Big Data analytics techniques and the newest HPC architectures to leverage each other clearly appears as a key catalyst factor for the convergence process.

9.3.2. Inria International Partners

9.3.2.1. DataCloud@Work

Title: DataCloud@Work.

International Partner:

- Polytechnic University of Bucharest (Romania), Computer Science Department, Nicolae Tapus and Valentin Cristea.

Duration: 4 years.

Start year: 2013. The status of IIP was established right after the end of our former *DataCloud@work* Associate Team (2010–2012).

URL: https://www.irisa.fr/kerdata/doku.php?id=cloud_at_work:start.

Description: Our research topics address the area of distributed data management for cloud services, focusing on autonomic storage. The goal is explore how to build an efficient, secure and reliable storage IaaS for data-intensive distributed applications running in cloud environments by enabling an autonomic behavior.

9.3.3. Informal International Partners

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

9.4. International Research Visitors

9.4.1. Visits of International Scientists

Guillaume Aupy (Vanderbilt University) visited the KerData team for one week (February 2016).

9.4.2. Visits to International Teams

9.4.2.1. Research Stays Abroad

CIC-IPN, Mexico:
Participants: Gabriel Antoniu, Alexandru Costan, Luis Eduardo Pineda Morales, Pierre Matri.

From October 31 to November 4, four members of our team visited the Informatics Research Centre of the National Polytechnic Institute (CIC-IPN for its acronym in Spanish) in Mexico City, Mexico.

The visit was a follow up to previous discussions held with the Network and Data Science Laboratory. The goal is to create a scientific collaboration on the grounds of cloud-based big data for smart cities, for which a proposal has been submitted in August to the program ECOS-NORD (Mexico-France). The visit included scientific presentations from both teams, a plenary talk from KerData to the IPN community, as well as discussions on future common research lines. Additionally, we held meetings with the partnering coordinator to talk about possible funding sources for students exchanges.

ANL, USA:
Participant: Nathanaël Cheriere.

Nathanaël Cheriere visited Matthieu Dorier and Rob Ross at ANL for 5.5 months, co-funded by the PUF NextGen project in the context of the Joint Laboratory for Extreme-Scale Computing (JLESC).

Vanderbilt University, USA:
Participant: Tien-Dat Phan.

Tien-Dat Phan visited(Guillaume Aupy, Padma Raghavan at Vanderbilt University for 2 months, funded by Vanderbilt University.

Technische Universitat Munchen and Huawei Research Center in Munich:
Participant: Ovidiu-Cristian Marcu.

Ovidiu-Cristian Marcu is doing an internship at Huawei in Munich, Germany for 4 months, starting October 2016. The goal is to create a framework to improve memory management for streaming systems.

National University of Singapore, Singapore:
Participant: Tien-Dat Phan.

Tien-Dat Phan is visiting NUS (Bingsheng He) for 3 months, co-funded by a Mobility grant from University Bretagne Loire (UBL) and NUS.

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events Organisation

10.1.1.1. General Chair, Scientific Chair

- Luc Bougé: Vice-Chair of the Steering Committee of the Euro-Par Series of conferences.

10.1.2. Scientific Events Selection

10.1.2.1. Chair of Conference Program Committees

- Gabriel Antoniu: Vice-Chair of the Program Committee of the ACM/IEEE CCGrid 2016 international conference (Hybrid and Mobile Clouds Tracks), Cartagena, May 2016.
- Alexandru Costan: Program Co-Chair of the ScienceCloud 2016 international workshop held in conjunction with HPDC 2016, Kyoto, June 2016.

10.1.2.2. Member of the Conference Program Committees

- Gabriel Antoniu: ACM HPDC 2016, IEEE Cluster 2016, PDSW-DISCS workshop (held in conjunction with ACM/IEEE SC16 conference), ARMS-CC 2016 workshop (held in conjunction with the PODC 2016 conference).
- Luc Bougé: Euro-Par 2016, IPDPS 2017, ICDE 2017, Euro-Par 2017, ISPDC 2017.
- Alexandru Costan: ACM/IEEE SC'16 BoF Applications Track, ACM/IEEE CCGrid 2016, IEEE BigData 2016, ICPP 2016, ARMS-CC 2016 workshop (held in conjunction with PODC 2016), FiCLOUD 2016, ScienceCloud 2016 workshop (held in conjunction with HPDC 2016).
- Shadi Ibrahim: IEEE Cluster 2016, IEEE/ACM CCGrid 2016, IEEE ICPADS 2016, IEEE CloudCom 2016, IEEE ICA3PP 2016, SCRAMBL 2016 (held in conjunction with Euro-Par 2016).

10.1.2.3. Reviewer

- Alexandru Costan: ACM/IEEE SC16, ACM HPDC 2016, IEEE Cluster 2016.
- Shadi Ibrahim: ACM HPDC 2016, Euro-Par 2016.

10.1.3. Journal

10.1.3.1. Member of the Editorial Boards

- Gabriel Antoniu: Future Generation Computer Systems, Special Issue on Resource Management for Big Data Platforms.
- Luc Bougé: Concurrency and Computation: Practice and Experience, Special Issues on the Euro-Par conference.
- Alexandru Costan: Soft Computing Journal, Special Issue on Autonomic Computing and Big Data Platforms

10.1.3.2. Reviewer, Reviewing Activities

- Alexandru Costan: IEEE Transactions on Parallel and Distributed Systems, Future Generation Computer Systems, Concurrency and Computation Practice and Experience, IEEE Communications, IEEE Transactions on Storage, Information Sciences
- Shadi Ibrahim: IEEE Transactions on Parallel and Distributed Systems, Future Generation Computer Systems, IEEE Transactions on Big Data, IEEE Transactions on Cloud Computing, Springer Parallel Computing, Computers and Electrical Engineering, Journal of Healthcare Engineering

10.1.4. Invited Talks

- Gabriel Antoniu:

- – BDEC 2016: Invited keynote talk at the *4th Big Data and Exascale Computing (BDEC) workshop*, Frankfurt, June 2016.
  – First Chinese-French Workshop on Extreme Computing: *Damaris: Jitter-Free I/O Management and In Situ Visualization of HPC Simulations using Dedicated Cores*, Guangzhou, May 2016.
  – 5th JLESC workshop: *Spark versus Flink: Understanding Performance in Big Data Analytics Frameworks*, Lyon, June 2016.
  – Inria/CIC-IPN workshop: *Scalable Big Data Processing on Clouds: A-Brain and Z-CloudFlow*, Mexico City, November 2016.
  – Inria/Technicolor workshop: *Spark versus Flink: Understanding Performance in Big Data Analytics Frameworks*, Rennes, November 2016.
  – 6th JLESC workshop: *Storage-Based Convergence Between HPC and Big Data*, Kobe, Japan, December 2016.
- Luc Bougé:
  – Comin Labs-DGA-ENSAI BigData day: *Support logiciel pour la gestion de données distribuées à très grande échelle*, IRISA, January 2016.
  – Société des agrégés: *Teaching informatics as a first-class subject*, annual meeting of the Regional Section, April 2016.
  – Luminy Algorithmics and Programming School: *Big Data: Tremendous challenges, great solutions*, Preparatory school teachers in Mathematics and Informatics, May 2016.
- Alexandru Costan:
  – UPB Scientific Days: *Big Data and Extreme Computing: A Storage-Based Pathway to Convergence*, The UPB Research Workshop on Distributed Systems, University Politehnica of Bucharest, June 2016.
  – Inria/CIC-IPN workshop: *Science Driven, Scalable Data-Intensive Processing on Clouds*, Mexico City, November 2016.
- Shadi Ibrahim:
  – Inria Scientific Days: *Big Data management at scale*, Rennes, June 2016.
- Chi Zhou:
  – 5th JLESC workshop: *Incorporating Probabilistic Optimizations for Resource Provisioning of Cloud Workflow Processing*, Lyon, June 2016.
- Nathanaël Cheriere:
  – 6th JLESC Workshop: *Accelerating the Scatter Operation on Dragonfly Networks*, Kobe, Japan, December 2016.
- Orçun Yildiz:
  – Grid'5000 Winter School: *Investigating the Root Causes of I/O Interference on Grid'5000*, Grenoble, February 2016.
  – 6th JLESC Workshop: *Towards Efficient Big Data Processing in HPC Systems*, Kobe, Japan, December 2016.
- Luis Eduardo Pineda Morales:
  – 5th JLESC workshop: *Exploring Elastic Scaling on Chameleon Cloud*, Lyon, June 2016.
  – Inria / CIC-IPN workshop: *Metadata Management for Geo-distributed Cloud Workflows*, Mexico City, November 2016.

10.1.5. Leadership within the Scientific Community

- Gabriel Antoniu: Scientific leader of the KerData project-team.

- Gabriel Antoniu: Topic leader for Inria for the *Data storage, I/O and in situ processing* topic, supervising collaboration activities in this area within the JLESC, Joint Inria-Illinois-ANL-BSC-JSC-RIKEN/AICS Laboratory for Extreme-Scale Computing.
- Luc Bougé: serves as a Vice-President of the *French Society for Informatics* (SIF), in charge of the teaching department.
- Gabriel Antoniu: Work package leader within the BigStorage H2020 ETN project for the *Data Science* work package.
- Alexandru Costan: Leader of the *Smart Cities* Working Group within the BigStorage H2020 ETN project.
- Shadi Ibrahim: Leader for the *Resource Management and Scheduling for Data-Intensive HPC Workflows* activity within the JLESC, Joint Inria-Illinois-ANL-BSC-JSC-RIKEN/AICS Laboratory for Extreme-Scale Computing.

10.1.6. Scientific Expertise

- Gabriel Antoniu served as a project evaluator for the ANR 2016 call (Phase 1).
- Luc Bougé: Member of the jury for the *Agrégation de mathématiques* and the *CAPES of mathématiques*. These national committees select high-school mathematics teachers in secondary schools and high-schools, respectively.
- Luc Bougé has been solicited by the Ministry of Education to participate to the committee in charge of designing and preparing the new *Informatics track* in the CAPES of mathematics. It will be offered for the 2017 session.
- Shadi Ibrahim served as a project evaluator in the DOE-ECP Program 2016: The research and development in Software Technology of the US Department of Energy's (DOE's) *Exascale Computing Project* (ECP).

10.1.7. Research Administration

- Luc Bougé: Nominated to seat at the CNU (*National University Council*) in the *Informatics* Section (27). His term ended in November 2016.

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

Gabriel Antoniu

- Master (Engineering Degree, 5th year): Big Data, 24 hours (lectures), M2 level, ENSAI (*École nationale supérieure de la statistique et de l'analyse de l'information*), Bruz, France.
- Master : Cloud Computing, 15 hours (lectures and lab sessions), M2 level, ENSAI (*École nationale supérieure de la statistique et de l'analyse de l'information*), Bruz, France.
- Master: Distributed Systems, 8 hours (lectures), M2 level, ALMA Master, Distributed Architectures module, University of Nantes, France.
- Master: Scalable Distributed Systems, 12 hours (lectures), M1 level, SDS Module, EIT ICT Labs Master School, France.

Luc Bougé

- Bachelor: Introduction to programming concepts, 36 hours (lectures), L3 level, Informatics program, ENS Rennes, France.
- Master: Introduction to compilation, 24 hours (exercice and practical classes), M1 level, Informatics program, Univ. Rennes I, France.

Alexandru Costan

- Bachelor: Software Engineering and Java Programming, 28 hours (lab sessions), L3, INSA Rennes.
- Bachelor: Databases, 68 hours (lectures and lab sessions), L2, INSA Rennes, France.
- Bachelor: Practical case studies, 24 hours (project), L3, INSA Rennes.
- Master: Big Data and Applications, 36h hours (lectures, lab sessions, project), M1, INSA Rennes.

Shadi Ibrahim

- Master (Engineering Degree, 5th year): Big Data, 24 hours (lectures and lab sessions), M2 level, ENSAI (*École nationale supérieure de la statistique et de l'analyse de l'information*), Bruz, France.
- Master : Cloud Computing and Hadoop Technologies, 16.5 hours (lectures and lab sessions), M2 level, ENSAI (*École nationale supérieure de la statistique et de l'analyse de l'information*), Bruz, France.
- Master: Cloud and Big data, 24 hours (lectures and lab sessions), M1 level, ENS Rennes, France.
- Master: Cloud1, Map-Reduce, (lectures, lab sessions), 15 hours (lectures and lab sessions), M2 level, Ecole des Mines de Nantes (EMN Nantes), Nantes, France.

10.2.2. Supervision

10.2.2.1. PhD in progress

Lokman Rahmani: *Big Data Management For Next Generation High Performance Computing Systems*, thesis started in October 2013, co-advised by Gabriel Antoniu and Luc Bougé.

Luis Eduardo Pineda Morales: *Efficient Big Data Management for Geographically Distributed Workflows*, thesis started in January 2014, co-advised by Alexandru Costan and Gabriel Antoniu. Defense planned in Spring 2017.

Orçun Yildiz: *Energy-Efficient Big Data Management in Petascale Supercomputers and Beyond*, thesis started in September 2014, co-advised by Shadi Ibrahim and Gabriel Antoniu.

Tien-Dat Phan: *Green Big Data Processing in Large-scale Clouds*, thesis started in October 2014, co-advised by Shadi Ibrahim and Luc Bougé.

Pierre Matri: *Predictive Models for Big Data*, thesis started in March 2015, co-advised by María Pérez and Gabriel Antoniu.

Mohammed-Yacine Taleb: *Energy-impact of data consistency management in Clouds and Beyond*, thesis started in August 2015, co-advised by Shadi Ibrahim and Gabriel Antoniu.

Ovidiu-Cristian Marcu: *Efficient data transfer and streaming strategies for workflow-based Big Data processing*, thesis started in October 2015, co-advised by Alexandru Costan and Gabriel Antoniu.

Nathanaël Cheriere: *Resource Management and Scheduling for Big Data Applications in Large-scale Systems*, thesis started in September 2016, co-advised by Shadi Ibrahim and Gabriel Antoniu.

Paul Le Noac'h: *Workflow Data Management as a Service for Multi-Site Applications*, thesis started in November 2016, co-advised by Alexandru Costan and Luc Bougé.

10.2.3. Juries

Gabriel Antoniu: Referee for the PhD thesis of Ms. Zhou Chi at the Nanyang Technological University (NTU), Singapore (January 2016).

Luc Bougé: Referee for the PhD thesis of Matthieu Perrin, LINA, Univ. Nantes (June 2016). Member of several PhD and HDR thesis juries in France.

10.2.4. Miscellaneous

10.2.4.1. Responsibilities

Luc Bougé: Co-ordinator between ENS Rennes and the Inria Research Center and the IRISA laboratory.

Luc Bougé: In charge of the Bachelor level (L3) and of the student seminar series at the Informatics Departement of ENS Rennes.

Alexandru Costan: In charge of communication at the Computer Science Department of INSA Rennes.

Alexandru Costan: In charge of the organization of the IRISA D1 Department Seminar.

Shadi Ibrahim: Member of Grid'5000 Sites Committee: Responsible for the Rennes site.

10.2.4.2. Tutorials

Gabriel Antoniu and Shadi Ibrahim gave tutorials on *Big Data technologies and Hadoop* at the BigStorage Winter School in Barcelona (March 2016).

Shadi Ibrahim gave a Tutorial on *Green Big Data Processing using Hadoop* at the at the Euro-Par 2016 conference, Grenoble, France (with Anne-Cécile Orgerie).

10.3. Popularization

Luc Bougé:

Master Program, Rennes. Invited presentation to the M2 students about *Informatics as a scientific activity: Toward a responsible research* (December 2016).

Alexandru Costan:

Master Program, Rennes. Invited presentation to the M2 students about *Big Data Analytics* (November 2016).

11. Bibliography

Major publications by the team in recent years

[1] N. CHERIERE, M. DORIER.*Design and Evaluation of Topology-aware Scatter and AllGather Algorithms for Dragonfly Networks*, November 2016, Supercomputing 2016, Poster, https://hal.inria.fr/hal-01400271.

[2] A. COSTAN, R. TUDORAN, G. ANTONIU, G. BRASCHE.*TomusBlobs: Scalable Data-intensive Processing on Azure Clouds*, in "CCPE - Concurrency and Computation: Practice and Experience", May 2013, https://hal.inria.fr/hal-00767034.

[3] B. DA MOTA, R. TUDORAN, A. COSTAN, G. VAROQUAUX, G. BRASCHE, P. J. CONROD, H. LEMAITRE, T. PAUS, M. RIETSCHEL, V. FROUIN, J.-B. POLINE, G. ANTONIU, B. THIRION.*Machine Learning Patterns for Neuroimaging-Genetic Studies in the Cloud*, in "Frontiers in Neuroinformatics", April 2014, vol. 8, https://hal.inria.fr/hal-01057325.

[4] M. DORIER, G. ANTONIU, F. CAPPELLO, M. SNIR, L. ORF.*Damaris: How to Efficiently Leverage Multicore Parallelism to Achieve Scalable, Jitter-free I/O*, in "CLUSTER - IEEE International Conference on Cluster Computing", Beijing, China, IEEE, September 2012, https://hal.inria.fr/hal-00715252.

[5] M. DORIER, G. ANTONIU, R. ROSS, D. KIMPE, S. IBRAHIM.*CALCioM: Mitigating I/O Interference in HPC Systems through Cross-Application Coordination*, in "IPDPS - International Parallel and Distributed Processing Symposium", Phoenix, United States, May 2014, https://hal.inria.fr/hal-00916091.

[6] M. DORIER, M. DREHER, T. PETERKA, G. ANTONIU, B. RAFFIN, J. M. WOZNIAK.*Lessons Learned from Building In Situ Coupling Frameworks*, in "ISAV 2015 - First Workshop on In Situ Infrastructures for Enabling Extreme-Scale Analysis and Visualization (held in conjunction with SC15)", Austin, United States, November 2015 [*DOI :* 10.1145/2828612.2828622], https://hal.inria.fr/hal-01224846.

[7] M. DORIER, S. IBRAHIM, G. ANTONIU, R. ROSS.*Omnisc'IO: A Grammar-Based Approach to Spatial and Temporal I/O Patterns Prediction*, in "SC14 - International Conference for High Performance Computing, Networking, Storage and Analysis", New Orleans, United States, IEEE, ACM, November 2014, https://hal.inria.fr/hal-01025670.

[8] M. DORIER, S. IBRAHIM, G. ANTONIU, R. ROSS.*Using Formal Grammars to Predict I/O Behaviors in HPC: the Omnisc'IO Approach*, in "TPDS - IEEE Transactions on Parallel and Distributed Systems", October 2015 [*DOI :* 10.1109/TPDS.2015.2485980], https://hal.inria.fr/hal-01238103.

[9] B. NICOLAE, G. ANTONIU, L. BOUGÉ, D. MOISE, A. CARPEN-AMARIE.*BlobSeer: Next-Generation Data Management for Large-Scale Infrastructures*, in "JPDC - Journal of Parallel and Distributed Computing", February 2011, vol. 71, n^{o} 2, p. 169–184, http://hal.inria.fr/inria-00511414/en/.

[10] B. NICOLAE, J. BRESNAHAN, K. KEAHEY, G. ANTONIU.*Going Back and Forth: Efficient Multi-Deployment and Multi-Snapshotting on Clouds*, in "HPDC 2011 - The 20th International ACM Symposium on High-Performance Parallel and Distributed Computing", San José, CA, United States, June 2011, http://hal.inria.fr/inria-00570682/en.

[11] R. TUDORAN, A. COSTAN, G. ANTONIU.*OverFlow: Multi-Site Aware Big Data Management for Scientific Workflows on Clouds*, in "IEEE Transactions on Cloud Computing", June 2015 [*DOI :* 10.1109/TCC.2015.2440254], https://hal.inria.fr/hal-01239128.

Publications of the year

Articles in International Peer-Reviewed Journal

[12] M. DORIER, G. ANTONIU, F. CAPPELLO, M. SNIR, R. R. SISNEROS, O. YILDIZ, S. IBRAHIM, T. PETERKA, L. G. ORF.*Damaris: Addressing Performance Variability in Data Management for Post-Petascale Simulations*, in "ACM Transactions on Parallel Computing", 2016, https://hal.inria.fr/hal-01353890.

[13] M. DORIER, S. IBRAHIM, G. ANTONIU, R. ROSS.*Using Formal Grammars to Predict I/O Behaviors in HPC: the Omnisc'IO Approach*, in "IEEE Transactions on Parallel and Distributed Systems", 2016 [*DOI :* 10.1109/TPDS.2015.2485980], https://hal.inria.fr/hal-01238103.

[14] M. DORIER, O. YILDIZ, S. IBRAHIM, A.-C. ORGERIE, G. ANTONIU.*On the energy footprint of I/O management in Exascale HPC systems*, in "Future Generation Computer Systems", March 2016, vol. 62, p. 17–28 [*DOI :* 10.1016/J.FUTURE.2016.03.002], https://hal.inria.fr/hal-01330735.

[15] S. IBRAHIM, T.-D. PHAN, A. CARPEN-AMARIE, H.-E. CHIHOUB, D. MOISE, G. ANTONIU.*Governing Energy Consumption in Hadoop through CPU Frequency Scaling: an Analysis*, in "Future Generation Computer Systems", January 2016, 14 [*DOI :* 10.1016/J.FUTURE.2015.01.005], https://hal.inria.fr/hal-01166252.

[16] Y. SIMMHAN, L. RAMAKRISHNAN, G. ANTONIU, C. GOBLE.*Cloud computing for data-driven science and engineering: Special issue on the Cloud computing for data-driven science and engineering workshop (ScienceCloud 2012)*, in "Concurrency and Computation: Practice and Experience", 2016, vol. 28, n^{o} 4, p. 947–949 [*DOI :* 10.1002/CPE.3668], https://hal.inria.fr/hal-01351218.

[17] R. TUDORAN, A. COSTAN, G. ANTONIU.*OverFlow: Multi-Site Aware Big Data Management for Scientific Workflows on Clouds*, in "IEEE Transactions on Cloud Computing", 2016 [*DOI :* 10.1109/TCC.2015.2440254], https://hal.inria.fr/hal-01239128.

[18] R. TUDORAN, A. COSTAN, O. NANO, I. SANTOS, H. SONCU, G. ANTONIU.*JetStream: Enabling high throughput live event streaming on multi-site clouds*, in "Future Generation Computer Systems", January 2016, vol. 54 [*DOI :* 10.1016/J.FUTURE.2015.01.016], https://hal.inria.fr/hal-01239124.

[19] S. WU, S. TAO, X. LING, H. FAN, H. JIN, S. IBRAHIM.*iShare: Balancing I/O performance isolation and disk I/O efficiency in virtualized environments*, in "Concurrency and Computation: Practice and Experience", 2016, https://hal.inria.fr/hal-01338404.

[20] O. YILDIZ, S. IBRAHIM, G. ANTONIU.*Enabling Fast Failure Recovery in Shared Hadoop Clusters: Towards Failure-Aware Scheduling*, in "Future Generation Computer Systems", March 2016 [*DOI :* 10.1016/J.FUTURE.2016.02.015], https://hal.inria.fr/hal-01338336.

International Conferences with Proceedings

[21] N. CHERIERE, P. DONAT-BOUILLUD, S. IBRAHIM, M. SIMONIN.*On the Usability of Shortest Remaining Time First Policy in Shared Hadoop Clusters*, in "SAC 2016-The 31st ACM/SIGAPP Symposium on Applied Computing", Pisa, Italy, April 2016, https://hal.inria.fr/hal-01239341.

[22] M. DORIER, R. R. SISNEROS, L. BAUTISTA-GOMEZ, T. PETERKA, L. ORF, L. RAHMANI, G. ANTONIU, L. BOUGÉ.*Adaptive Performance-Constrained In Situ Visualization of Atmospheric Simulations*, in "Cluster 2016 - The IEEE 2016 International Conference on Cluster Computing", Taipei, Taiwan, 2016, https://hal.inria.fr/hal-01351919.

[23] O.-C. MARCU, A. COSTAN, G. ANTONIU, M. S. PÉREZ.*Spark versus Flink: Understanding Performance in Big Data Analytics Frameworks*, in "Cluster 2016 - The IEEE 2016 International Conference on Cluster Computing", Taipei, Taiwan, September 2016, https://hal.inria.fr/hal-01347638.

[24] P. MATRI, A. COSTAN, G. ANTONIU, J. MONTES, M. S. PÉREZ.*Towards Efficient Location and Placement of Dynamic Replicas for Geo-Distributed Data Stores*, in "ScienceCloud'16 - 7th Workshop on Scientific Cloud Computing (in conjunction with ACM HPDC 2016)", Kyoto, Japan, June 2016 [*DOI :* 10.1145/2913712.2913715], https://hal.inria.fr/hal-01304328.

[25] *Best Paper*
P. MATRI, A. COSTAN, G. ANTONIU, J. MONTES, M. S. PÉREZ.*Tyr: Blob Storage Meets Built-In Transactions*, in "IEEE ACM SC16 - The International Conference for High Performance Computing, Networking, Storage and Analysis 2016", Salt Lake City, United States, November 2016, https://hal.inria.fr/hal-01347652.

[26] L. PINEDA-MORALES, J. LIU, A. COSTAN, E. PACITTI, G. ANTONIU, P. VALDURIEZ, M. MATTOSO.*Managing Hot Metadata for Scientific Workflows on Multisite Clouds*, in "BIGDATA 2016 - 2016 IEEE International Conference on Big Data", Washington, United States, December 2016, https://hal.inria.fr/hal-01395715.

[27] O. YILDIZ, M. DORIER, S. IBRAHIM, R. ROSS, G. ANTONIU.*On the Root Causes of Cross-Application I/O Interference in HPC Storage Systems*, in "IPDPS 2016 - The 30th IEEE International Parallel and Distributed Processing Symposium", Chicago, United States, May 2016, https://hal.inria.fr/hal-01270630.

Scientific Books (or Scientific Book chapters)

[28] B. MEMISHI, S. IBRAHIM, M. S. PÉREZ-HERNÁNDEZ, G. ANTONIU.*On the Dynamic Shifting of the MapReduce Timeout*, in "Managing and Processing Big Data in Cloud Computing", R. KANNAN, R. U. RASOOL, H. JIN, S. BALASUNDARAM (editors), IGI Global, 2016, https://hal.inria.fr/hal-01338393.

[29] A. C. ZHOU, B. HE, S. IBRAHIM.*A Taxonomy and Survey of Scientific Computing in the Cloud*, in "Big Data: Principles and Paradigms", eScience and Big Data Workflows in Clouds: A Taxonomy and Survey, Morgan Kaufmann, June 2016, https://hal.inria.fr/hal-01346745.

Research Reports

[30] M. DORIER, R. R. SISNEROS, L. BAUTISTA-GOMEZ, T. PETERKA, L. G. ORF, R. ROSS, L. RAHMANI, G. ANTONIU, L. BOUGÉ.*Performance-Constrained In Situ Visualization of Atmospheric Simulations*, Inria Rennes - Bretagne Atlantique, February 2016, n^{o} RR-8855, 27, https://hal.inria.fr/hal-01273718.

[31] P. MATRI, A. COSTAN, G. ANTONIU, J. MONTES, M. S. PÉREZ.*Týr: Efficient Transactional Storage for Data-Intensive Applications*, Inria Rennes Bretagne Atlantique ; Universidad Politécnica de Madrid, January 2016, n^{o} RT-0473, 25, https://hal.inria.fr/hal-01256563.

Other Publications

[32] N. CHERIERE, M. DORIER.*Design and Evaluation of Topology-aware Scatter and AllGather Algorithms for Dragonfly Networks*, November 2016, Supercomputing 2016, Poster, https://hal.inria.fr/hal-01400271.

[33] L. RAHMANI, M. DORIER, L. BOUGÉ, G. ANTONIU, R. R. SISNEROS, T. PETERKA.*Towards Smart Visualization Framework for Climate Simulations*, March 2016, working paper or preprint, https://hal.inria.fr/hal-01290268.

[34] Y. TALEB, S. IBRAHIM, G. ANTONIU, T. CORTES.*Understanding how the network impacts performance and energy-efficiency in the RAMCloud storage system*, October 2016, working paper or preprint, https://hal.inria.fr/hal-01376923.

References in notes

[35] *Amazon Elastic Map-Reduce (EMR)*, https://aws.amazon.com/emr/.

[36] *The Decaf Project*, https://bitbucket.org/tpeterka1/decaf.

[37] *Digital Single Market*, 2015, https://ec.europa.eu/digital-single-market/en/digital-single-market.

[38] *European Exascale Software Initiative*, 2013, http://www.eesi-project.eu.

[39] *The European Technology Platform for High-Performance Computing*, 2012, http://www.etp4hpc.eu.

[40] *European Cloud Strategy*, 2012, https://ec.europa.eu/digital-single-market/en/european-cloud-computing-strategy.

[41] *Apache Flink*, 2016, http://flink.apache.org.

[42] *International Exascale Software Program*, 2011, http://www.exascale.org/iesp/Main_Page.

[43] *Scientific challenges of the Inria Rennes-Bretagne Atlantique research centre*, 2016, https://www.inria.fr/en/centre/rennes/research.

[44] *Inria's strategic plan "Towards Inria 2020"*, 2016, https://www.inria.fr/en/institute/strategy/strategic-plan.

[45] *Joint Laboratory for Extreme Scale Computing (JLESC)*, https://jlesc.github.io.

[46] *Apache Spark*, http://spark.apache.org.

[47] *Storm*, http://storm-project.net/.

[48] *The FlowVR Project*, 2014, http://flowvr.sourceforge.net/.

[49] T. Akidau, A. Balikov, K. Bekiroğlu, S. Chernyak, J. Haberman, R. Lax, S. McVeety, D. Mills, P. Nordstrom, S. Whittle.*MillWheel: fault-tolerant stream processing at internet scale*, in "Proceedings of the VLDB Endowment", 2013, vol. 6, n^{o} 11, p. 1033–1044.

[50] J. Dean, S. Ghemawat.*MapReduce: simplified data processing on large clusters*, in "Communications of the ACM", 2008, vol. 51, n^{o} 1, p. 107–113.

[51] S. Wilde, M. Hategan, J. M. Wozniak, B. Clifford, D. Katz, I. T. Foster.*Swift: A language for distributed parallel scripting*, in "Parallel Computing", 2011, vol. 37, n^{o} 9, p. 633–652, http://dx.doi.org/10.1016/j.parco.2011.05.005.

Team LACODAM

Large Scale Collaborative Data Mining

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

RESEARCH CENTER
Rennes - Bretagne-Atlantique

THEME
Data and Knowledge Representation and Processing

Table of contents

Team LACODAM

Creation of the Team: 2016 January 01

Keywords:

Computer Science and Digital Science:

3. - Data and knowledge
3.1.1. - Modeling, representation
3.2. - Knowledge
3.2.2. - Knowledge extraction, cleaning
3.2.3. - Inference
3.3. - Data and knowledge analysis
3.3.1. - On-line analytical processing
3.3.2. - Data mining
3.3.3. - Big data analysis
3.4. - Machine learning and statistics
3.4.1. - Supervised learning
4.9.1. - Intrusion detection
7.1. - Parallel and distributed algorithms
8. - Artificial intelligence
8.1. - Knowledge
8.2. - Machine learning
8.6. - Decision support

Other Research Topics and Application Domains:

1.2. - Ecology
2.4. - Therapies
2.4.2. - Drug resistance
3.4.3. - Pollution
3.5. - Agronomy
4. - Energy
5.4. - Microelectronics
6.2. - Network technologies
9. - Society and Knowledge

1. Members

Research Scientists

René Quiniou [Inria, Researcher]
Torsten Schaub [University of Potsdam, Germany, International senior chair, senior Researcher, HDR]

Faculty Members

Alexandre Termier [Team leader, Univ. Rennes I, Faculty Member, Professor, HDR]
Marie-Odile Cordier [Univ. Rennes I, Faculty Member, Emeritus Professor, HDR]
Véronique Masson [Univ. Rennes I, Associate Professor]
Laurence Rozé [INSA Rennes, Associate Professor]

Thomas Guyet [Agrocampus Ouest Rennes, Associate Professor, Inria half-secondment from Sept. 2015 to Aug. 2016]
Christine Largouët [Agrocampus Ouest Rennes, Associate Professor]

Technical Staff

Louis Bonneau de Beaufort [Agrocampus Ouest Rennes]

PhD Students

Yann Dauxais [Univ. Rennes I]
Clément Gautrais [Univ. Rennes I]
Maël Guillemé [Energiency, granted by CIFRE, from Oct 2016]

Post-Doctoral Fellows

Benjamin Négrevergne [Inria, until Sep 2016]
Ahmed Samet [Univ. Rennes I, from May 2016]

Administrative Assistant

Marie-Noëlle Georgeault [Inria]

Others

Philippe Besnard [CNRS, Research Scientist, HDR]
Anne-Isabelle Graux [INRA, Researcher]

2. Overall Objectives

2.1. Overall Objectives

Ubiquitous data collection is providing our society with tremendous volumes of data about human, environmental and industrial activity. These ever increasing volumes of collected data hold the keys to new discoveries, both in the industrial and scientific domains. However, those keys will only be accessible to those who can make sense of such data. Making sense of data is a hard problem, requiring a good understanding of the data at hand, of the many data analysis tools and methods, and a good capacity to infer knowledge from the results of such tools. Such skills have been grouped under the umbrella term "Data Science" and lots of efforts are being done on education and research in this area. "Data Scientist" is currently the most sought after job in the US, as the demand far exceeds the number of competent professionals. Nowadays, the main problem of data science is that despite considerable improvements, it is still mostly a "manual" process: current data analysis tools still require an important human effort and know-how, making data analysis a lengthy, partial and error-prone process. This is true even for data science experts, and current approaches are mostly out of reach of non-specialists.

We claim that nowadays, Data Science is in its "Iron Age": good tools are available, however skilled craftsmen are required to use them in order to transform raw material (the data) into finished products (knowledge, decisions). We foresee that in a decade from now, we should be in an "Industrial Age" of Data Science, where more elaborate tools will alleviate a lot of the human work required in Data Science. Basic Data Science tasks will no longer require a skilled data scientist, but software tools will enable small companies or even individuals to get valuable knowledge from their data, which is not possible currently. Skilled data scientists will thus be fully available to work on the hard tasks that matter, with a drastic productivity improvement thanks to better tools doing the tedious work for them.

The objective of the Lacodam team is to considerably facilitate the process of making sense from large quantities of data, either to derive new knowledge or for making better decisions. Nowadays, this process is mostly manual, and relies on the analyst's understanding of the domain, of the data at hand and of a plethora of complex computational tools. We envision a novel generation of data analysis and decision support tools that require significantly less tedious human work, relying only on few interactions with high added value. The solutions we foresee requires to bridge data mining techniques with artificial intelligence (AI) approaches, both to take knowledge into account in a principled way, and to introduce automated reasoning techniques in knowledge discovery workflows. Such solutions can be seen as "second order" AI tasks: they exploit AI techniques (for example, planning) in order to pilot more classical AI tasks such as data mining and decision support.

3. Research Program

3.1. Introduction

The three research axes of the Lacodam project-team are the following. First, we briefly introduce these axes, as well as their interplay:

- The first research axis is dedicated to the design of *novel pattern mining methods*. Pattern mining is one of the most important approaches to discover novel knowledge in data, and one of our strongest areas of expertise. Work in this axis will be the most fundamental of all three axes, and is expected to serve as foundations for work on the other two axes.
- The second axis tackles another aspect of knowledge discovery in data: the *interaction between the user and the system*, in order to co-discover novel knowledge. Our team has a long experience to collaborate with domain experts, and is thus especially aware of the need to improve such interaction.
- The third axis concerns *decision support*. With the help of methods from the two previous axes, our goal here is to design systems that can either help humans to take better decisions in precise applicative contexts, or to allow machines to automatically take relevant decisions in situations where extremely fast reaction time is required.

The following figure sums up the detailed work presented in the next few pages: on the sides are the three research axes of the team (X-axis) and our main applications areas (Y-axis). In the middle are colored squares that represent the precise research topics of the team that will be described in this section, placed relatively to their axis and main application area. Lines represent projects that can link several topics, and that are also connected to their main application area.

3.2. Pattern mining algorithms

Twenty years of research in pattern mining have resulted in efficient approaches to handle the algorithmic complexity of the problem. Existing algorithms are now able to efficiently extract patterns with complex structures (ex: sequences, graphs, co-variations) from large datasets. However, when dealing with large, real world datasets, these methods still output a huge set of patterns, which is impractical for human analysis. This problem is called pattern explosion. The ongoing challenge of pattern mining research is to extract fewer but more meaningful patterns. The Lacodam team is committed to solve the pattern explosion problem following four research topics:

- the design of dedicated algorithms for mining temporal patterns
- the design of flexible pattern mining approaches
- the selection of interesting data mining results
- the design of parallel pattern algorithms to ensure scalability

The originality of our contributions relies on the exploration of knowledge-based approaches whose principle is to incorporate dedicated domain knowledge (aka application background knowledge) deep into the mining process. While most of the data mining approaches are based on agnostic approaches that are designed to cope with the pattern explosion, we propose to develop data mining techniques relying on knowledge-based artificial intelligence techniques. This covers the use of structured knowledge representations, as well as reasoning methods, in combination with mining.

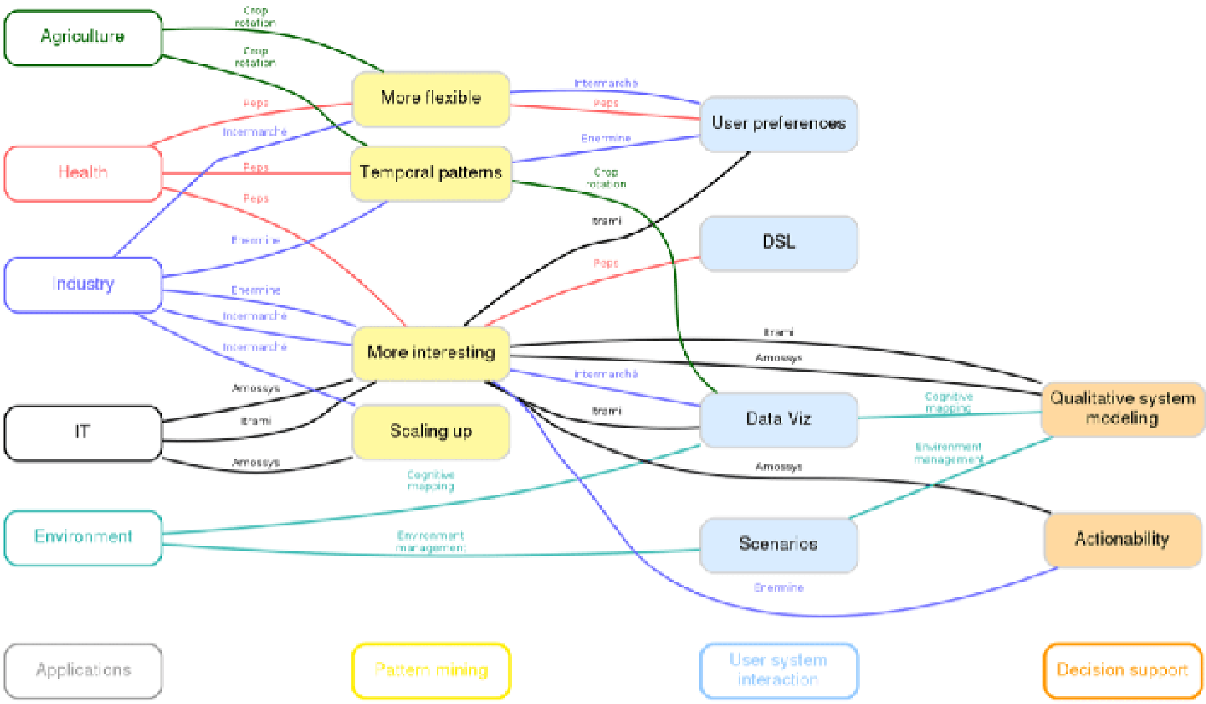


Figure 1. Lacodam research topics organized by axis and application

The first approach concerns the classical approach of pattern mining which consists in using expert knowledge to define new pattern types (and related algorithms) that can solve applicative issues. In particular, we investigate how to handle temporality in pattern representations which turns out to be important in many real world applications (in particular for decision support) and deserves particular attention.

The two other approaches aim at proposing alternative pattern mining methods to let the user incorporate, by her own, knowledge that will help define her pattern domain of interest. Flexible pattern mining approaches enable analysts to easily incorporate extra knowledge, for example domain related constraints, in order to extract only the most relevant patterns. On the other hand, the selection of interesting data mining results aims at devising strategies to filter out the results that are useless for the data analyst. Beside the challenge related to algorithmic efficiency of such approaches, we are interested in formalizing the foundations of interestingness, according to background knowledge modeled with logic knowledge representation paradigms.

Last, pattern mining algorithms are computation-intensive, it is thus important to exploit all the available computing power. Parallelism is for a foreseeable future one of the main ways to speed up computations, and we have a strong competence on the design of parallel pattern mining algorithms. We will exploit this competence in order to guarantee that our approaches scale up to the real data provided by our partners.

3.3. User/system interaction

As we pointed out before, there is a strong need to present relevant patterns to the user. This can be done by using more specific constraints, background knowledge and/or tailor-made optimization functions. Due to the difficulty of determining these elements beforehand, one of the most promising solutions is that the system and the user co-construct the definition of most relevant patterns, i.e., to have a human in the loop. This requires to have means to present intermediate results to the user, and to get user feedback in order to guide the search space exploration process in the right direction. This is an important research axis for Lacodam, which will be tackled in several complementary ways:

- Domain Specific Languages: one way to interact with the user is to propose a Domain Specific Language (DSL) tailored to the domain at hand and to the analysis tasks to perform. The challenge is to propose a DSL allowing the users to easily express the required processing workflows, to deploy those workflows for mining large volumes of data and to offer as much automation as possible.
- What if / What for scenarios: we are also investigating the use of scenarios to query results from data mining processes, as well as other complex processes such as complex system simulations or model predictions. Such scenarios are answers to questions of the type “what if [situation]” or “what [should be done] for [expected outcome]”.
- User preferences: in exploratory analysis, users often do not have a precise enough idea of what they want, and are not able to formulate such queries. Lacodam is thus investigating simple ways for letting users express their interests and preferences, either during the mining process to guide the search space exploration, or after, to help in getting the most relevant results.
- Data visualization: most of the research directions presented in this document require users to examine patterns at some point. The output of most pattern mining algorithms is simply a (long) list of patterns. While this presentation can be sufficient in some applications, it is often not enough to provide a complete understanding, especially for non-experts in pattern mining. A transversal research topic that we want to develop in Lacodam is to propose data visualization techniques adequate to understanding output results. Numerous (failed) experiments have shown that data mining and data visualization are fields which require distinct skills, where researchers in one field usually do not make significant advances in the other field (this is detailed in [Keim 2010]). Thus, our strategy is to establish collaborations with prominent data visualization teams for this line of research, with a long term goal to recruit a specialist in data visualization if the opportunity arises.

3.4. Decision support

Patterns, especially predictive sequential patterns, resulting from mining a dataset have often a direct application in diagnosis. Lacodam inherits from the former Dream team a strong background in decision support systems, with an internationally recognized expertise in diagnosis. This AI subfield is concerned with determining if a system is operating normally or not, and if the system is in an abnormal state, to determine the cause of the faulty behavior. The considered system can as well be an agro- or eco-system, a software system or an animal or human being, as well.

The increasing volumes of data coming from a wide range of different systems (ex: sensor data from agro-environmental systems, log data from software systems, biological data coming from health monitoring systems) show that it is possible to gather more and more observations for such systems. Thus, it should be possible to exploit such observations to help human or software agents to take better decisions. Hence, while keeping the strong interest on decision support (and especially diagnosis) that existed in Dream, Lacodam adds the idea that the decision support systems should take advantage of the huge volumes of data available. This third and last research axis is thus a meeting point for all members of the team, as it requires to integrate AI techniques of traditional decision support systems with results from data mining techniques.

Two main research axes are investigated in Lacodam:

- Diagnosis-based approaches. We are exploring how to integrate knowledge found from pattern mining approaches, possibly with the help of interactive methods, into the qualitative models. The goal of such work is to partly automate the construction of the model, which can require a lot of human effort otherwise.
- Actionable patterns and rules. In many settings of "exploratory data mining", the actual interest of a pattern is hard to assess, as it may be hard to measure or may be subjective (resulting from introducing the user in the mining process). However, there exist applications where once patterns are found, there are well defined measures to define what this pattern will bring to the user. Further, patterns and rules that can lead to actual actions beneficial to the user are called actionable patterns. Such actionable patterns and rules are especially important for industry.

3.5. Long-term goals

The following perspectives are at the convergence of the three research axes presented before, and can be seen as the ideal towards which our efforts tend:

- Automating data science workflow discovery. The current methods for extracting knowledge from data and building decision support systems require a lot of human effort. Our three research axes aim at alleviating this effort, by devising methods that are more generic and by improving the interaction between the user and the system. An ideal solution would be that the user could forget completely about the existence of pattern mining or decision support methods. Instead the user would only loosely specify her problem, while the system would construct for her various data science / decision support workflows, possibly further refined via interactions.

  We consider that this is a second order AI task, where AI techniques such as planning are used to explore the workflow search space, the workflow itself being composed of data mining and/or decision support components. This is a strategic evolution for data science endeavors, were the demand far exceeds the available human skilled manpower.
- Logic argumentation based on epistemic interest. Having increasingly automated approaches will require better and better ways to handle the interactions with the user. Our second long term goal is to explore the use of logic argumentation as an interaction tool between users and a data analysis tool. Alongside visualization and interactive data mining tools, it can be a way for users to query in an intuitive manner both the results and the way they were obtained. Such querying can also help the expert to reformulate her query in an interactive analysis setting.

This research direction continues the work on “epistemic interest” presented before. Its goal is to exploit principles of interactive data analysis in the context of epistemic interest measures. Logic argumentation [Besnard 2014] can be a natural tool for interactions between the user and the system: display of possibly exhaustive list of arguments, relationships – whether reinforcement, compatibility or conflict – between arguments, variable degrees of arguments, and possible solutions for argument conflicts.

The first step is to define a formal argumentation framework for explaining data mining results. This implies to continue theoretical work on the foundations of argumentation in order to identify the most adapted framework (either existing or a new one to be defined). Logic argumentation may be implemented and deeply explored in ASP, allowing us to build on our expertise in this logic language.

- Collaborative feedback and knowledge management. We are convinced that improving the data science process, and possibly automating it, will rely at some point in the near future on the vast feedback that can be obtained by communities of user seamlessly collaborating over the web. Consider for example what has been achieved by collaborative platforms such as StackOverflow: it has become the reference site for any programming question.

  Data science is a more complex problem than programming, as in order to get help from the community, the user has to share her data and workflow, or at least some parts of them. This raises obvious privacy issues that may prevent this idea to succeed. As our research on automating the production of data science workflows should enable more people to have access to data science results, we are interested to investigate the design of collaborative platforms to exchange expert advices over data, workflows and analysis results, with an aim at exploiting this human feedback to improve the automated system with machine learning.

4. Application Domains

4.1. Introduction

The current period is extremely favorable for teams working in Data Science and Artificial Intelligence, and Lacodam is no exception. We are eager to see our work applied in real world applications, and have thus an important activity in maintaining strong ties with industrials partners concerned with marketing and energy as well as public partners working in health, agriculture and environment.

4.2. Industry

We present below our industrial collaborations. Some are well established partnerships, while others are more recent collaborations with local industries that wish to reinforce their Data Science R&D with us (e.g. STMicroelectronics, Energiency, Amossys).

- **Execution trace analysis for SOC debugging (STMicroelectronics)**. We have an ongoing collaborations with STMicroelectronics, which is one of the world top-5 electronic chip makers. Nowadays, set-top boxes, smartphones or onboard car computers are powered by highly integrated chips called System-on-Chip (SoC). Such chips contain on a single die processing units, memories, IO units and specialized accelerators (such as audio and video encoding/decoding). Programming SoC is a hard task due to their inherent parallelism, leading to subtle bugs when several components do not deliver their results within a given time frame. Existing debuggers and profilers are ill-adapted in this case because of their high intrusivity that modifies the timings. Hence the most used technique is to capture a trace of the execution and analyze it post-mortem. While Alexandre Termier was in Grenoble he initiated several works for analyzing such traces with pattern mining techniques, which he is now pursuing with his colleagues of the Lacodam project-team.

- **Resource consumption analysis for optimizing energy consumption and practices in industrial factories (Energiency)**. In order to increase their benefits, companies introduce more and more sensors in their factories. Thus, the resource (electricity, water, etc.) consumption of engines, workshops or factories are recorded in the form of times series or temporal sequences. The person who is in charge of resource consumption optimization needs better software than classical spreadsheets. He/she needs effective decision-aiding tools with statistical and artificial intelligence knowledge. The start-up Energiency aims at designing and offering such pieces of software for analyzing energy consumption. The starting CIFRE PhD thesis of Maël Guillemé aims at proposing new approaches and solutions from the data mining field to tackle this issue.
- **Security (Amossys)**. Current networks are faced with an increasing variety of attacks, from the classic « DDoS » that makes a server unusuable for a few fours, to advanced attacks that silently infiltrate a network and exfiltrate sensitive information monthes or even years later. Such intrusions, called APT (Advanced Persistent Threat) are extremely hard to detect, and this will become even harder as most communications will be encrypted. A promising solution is to work on "behavioral analysis", by discovering patterns based on the metadata of IP-packets. Such patterns can relate to an unusual sequencing of events, or to an unusual communication graph. Finding such complex patterns over a large volume of streaming data requires to revisit existing stream mining algorithms to dramatically improve their throughput, while guaranteeing a manageable false positive rate. We are collaborating on this topic with the Amossys company and the Emsec team of Irisa through the co-supervision of a CIFRE PhD (located in the Emsec team). Our goal is to design novel anomaly detection methods that can detect APT, and that scales on real traffic volumes.
- **Market basket data analysis (Intermarché) and multi-channel interaction data analysis (EDF) for better Customer Relationship Management (CRM)**. An important application domain of data mining for companies that deal with large numbers of customers is to analyze customer interaction data, either for marketing purposes or to improve the quality of service. We have activities in both settings. In the first case, we collaborate with a major french retailer, Intermarché, in order to detect customer churn by analyzing market basket data. In the second case, we collaborate with the major french power supplier, EDF, to discover actionable patterns for CRM aiming at avoiding reaching undesirable situations from logs of user interactions with the company (web clicks, phone calls, etc.).

4.3. Health

- **Care pathways analysis for supporting pharmaco-epidemiological studies**. Pharmaco-epidemiology applies the methodologies developed in general epidemiology to answer to questions about the uses and effects of health products, drugs [20], [19] or medical devices [17], on population. In classical pharmaco-epidemiology studies, people who share common characteristics are recruited to build a dedicated prospective cohort. Then, meaningful data (drug exposures, diseases, etc.) are collected from the cohort within a defined period of time. Finally, a statistical analysis highlights the links (or the lack of links) between drug exposures and outcomes (*e.g.* adverse effects). The main drawback of prospective cohort studies is the time required to collect the data and to integrate it. Indeed, in some cases of health product safety, health authorities have to answer quickly to pharmaco-epidemiology questions.

  New approaches of pharmaco-epidemiology consist in using large EHR (Electronic Health Records) databases to investigate the effects and uses (or misuses) of drugs in real conditions. The objective is to benefit from nationwide available data to answer accurately and in a short time pharmaco-epidemiological queries for national public health institutions. Despite the potential availability of the data, their size and complexity make their analysis long and tremendous. The challenge we tackle is the conception of a generic digital toolbox to support the efficient design of a broad range of pharmaco-epidemiology studies from EHR databases.

We propose to use pattern mining algorithm and reasoning techniques to analyse the typical care pathways of specific groups of patients.

To be able to answer the broad range of pharmaco-epidemiological queries from national public health institutions, the PEPS [0] platform exploits, in secondary use, the French health cross-schemes insurance system, called SNIIRAM. The SNIIRAM covers most of the French population with a sliding period of 3 past years. The main characteristics of this data warehouse are described in [18]. Contrary to local hospital EHR or even with other national initiatives, the SNIIRAM data warehouse covers a huge population. It makes possible studies on unfrequent drugs or diseases in real conditions of use. To tackle the volume and the diversity of the SNIIRAM data warehouse, a research program has been established to design an innovative toolbox. This research program is focused first on the modeling of care pathways from the SNIIRAM database and, second, on the design of tools supporting meaningful insights extraction about massive and complex care pathways by clinicians. In such database a care pathway is an individual sequence of drugs exposures, medical procedures and hospitalizations.

4.4. Agriculture and environment

- **Dairy farming**. The use and analysis of data acquired in dairy farming is a challenge both for data science and for animal science. Its goal is to improve farming conditions (health, welfare and environment) as well as farmers' income. Nowadays, animals are monitored by multiple sensors giving a wealth of heterogeneous data (ex: temperature, weight, milk composition…). Current techniques used by animal scientists focus mostly on mono-sensor approaches. The dynamic combination of several sensors could provide new services and information useful for dairy farming. A PhD thesis will begin soon to study such combinations of sensors and to investigate data mining methods, especially pattern mining algorithms. The challenge is to design new algorithms taking into account the data heterogeneity, coming both from their nature and the different time scales involved, and to produce patterns that are actually useful for dairy farming. This thesis will be an original and important contribution to the new challenge of the IoT (Internet of Things) and will interest domain actors to find new added value to a global data analysis. The PhD thesis will take place in an interdisciplinary setting bringing together computer scientists from Inria and animal scientists from INRA, both located in Rennes.

  Similar problems are investigated with the veterinary department of the University of Calgary in the context of cattle monitoring from multiple sensors placed on calves for the early detection of diseases.
- **Optimizing the nutrition of individual sow**. Another direction for further research is to combine data flow with prediction models in order to learn nutrition strategies. We are currently starting a project with INRAon the nutritional requirements and the optimal diet to be supplied to individual lactating sow. The research issue will be to develop decision algorithms for the determination of the optimal ration (amount and composition) to be fed to a given sow, on a given day, considering all the information available (real-time observation data flow and historical data). Issues concern the design of an incremental learning algorithm that will compute the animal profile and how to determine the best feeding plan. Efficiency issues of developed algorithms will also be considered since the proposed software should work in real-time on the automated feeder.
- **Ecosystem modeling and management**. Ongoing research on ecosystem management includes modelling of ecosystems and anthroprogenic pressures, with a special concern on the representation of socio-economical factors that impact human decisions. A main research issue is how to to represent these factors and how to integrate their impact on the ecosystem simulation model. This work is an ongoing cooperation with ecologists from the Marine Spatial Ecology of Queensland University, Australia and from Agrocampus Ouest.

[0] PEPS: Pharmaco-Epidémiologie et Produits de Santé – Pharmacoepidemiology of health products

5. Highlights of the Year

5.1. Highlights of the Year

- This year, we are extremely proud to have a total of 4 papers accepted at the IJCAI conference, the rank A+ conference on Artificial Intelligence.
- Another highlight of this year is that following the end of the former team, namely Dream, we could propose in 2016 a new team project, namely Lacodam, and follow smoothly all steps of the Inria project-team creation protocol. While the team is not officially created as of December 2016, our project has been positively evaluated both by Inria members and by international experts, and is thus likely to be created in early 2017.

6. New Software and Platforms

6.1. EcoMata

FUNCTIONAL DESCRIPTION

The EcoMata tool-box provides means for qualitative modeling and ecosystem exploration and for assisting the design of environmental guidelines. We have proposed a new qualitative approach for ecosystem modeling based on timed automata (TA) formalism combined to a high-level query language for exploring scenarios.

- Participants: Marie-Odile Cordier, Yulong Zhao, Christine Largouët and Thomas Guyet
- Contact: Christine Largouët

6.2. PaturMata

KEYWORDS: Bioinformatics - Biology
SCIENTIFIC DESCRIPTION

The Paturmata tool-box provides means for qualitative modeling and exploring agrosystems, specifically management of herd based on pasture. The system is modelled using a hierarchical hybrid model described in timed automata formalism.
FUNCTIONAL DESCRIPTION In the PaturMata software, users can create a pasture system description by entering herds and plots information. For each herd, the only parameter is the number of animals. For each plot, users should enter the surface, the density, the herb height, the distance to the milking shed, a herb growth profile and an accessibility degree.

Users then specify pasturing and fertilization strategies. Finally, users can launch a pasture execution. PaturMata displays the results and a detailed trace of pasture. Users can launch a batch of different strategies and compare the results in order to find the best pasture strategy.

PaturMata is developed in Java (Swing for the GUI) and the model-checker that is called for the timed properties verification is UPPAAL .

- Participants: Christine Largouët, Marie-Odile Cordier, Yulong Zhao
- Contact: Christine Largouët

6.3. QTempIntMiner

KEYWORDS: Data mining - Health - Medical - Physiology - Temporal information
SCIENTIFIC DESCRIPTION

QTempIntMiner: the QTempIntMiner (Quantitative Temporal Interval Miner) data mining software implements several algorithms presented in [46] and [3] (QTIAPriori and QTIPrefixSpan). The software is mainly implemented in Matlab. It uses the Mixmod toolbox [33] to compute multi-dimensional Gaussian distributions. The main features of QTempIntMiner are:

- a tool for generating synthetic noisy sequences of temporal events,
- an implementation of the QTempIntMiner , QTIAPriori and QTIPrefixSpan algorithms,
- a graphical interface that enables the user to generate or import data set and to define the parameters of the algorithm and that displays the extracted temporal patterns.
- a sequence transformer to process long sequences of temporal events. Long sequences are transformed into a database of short temporal sequences that are used as input instances for the available algorithms.

The software includes one new algorithm based on the separation of the set of intervals to extract more efficiently but less accurately the time interval in temporal patterns. This new algorithm version is still under evaluation on simulated and real datasets (care pathways).

- Participants: Thomas Guyet and René Quiniou
- Partner: AGROCAMPUS
- Contact: Thomas Guyet
- URL: http://people.irisa.fr/Thomas.Guyet/QTempIntMiner/

7. New Results

7.1. Introduction

In this section, we organize our contributions over three main research topics:

- Mining different kinds of patterns, from 7.2 to 7.9
- Data mining and decision support with ASP, from 7.10 to 7.13
- Model-based diagnosis, from 7.13 to 7.15.

7.2. Customer Purchase Signatures: a New Model in Grocery Retail Context

Participants: Clément Gautrais, Peggy Cellier [Lis], Thomas Guyet, René Quiniou, Alexandre Termier.

In the retail context, there is an increasing need for understanding individual customer behavior in order to personalize marketing actions. We propose the novel concept of customer signature, that identifies a set of important products that the customer refills regularly. Both the set of products and the refilling time periods give new insights on the customer behavior. Our approach is inspired by methods from the domain of sequence segmentation, thus benefiting from efficient exact and approximate algorithms. Experiments on a real massive retail dataset show the interest of the signatures for understanding individual customers (under submission to PAKDD 2017 conference).

This new model is used to detect and explain customer defection in a grocery retail context from the evolution of each customer basket content. It therefore provides actionable knowledge for the retailer at an individual scale. In addition, this model is able to identify customers that are likely to defect in the future months [16].

7.3. Discriminant Chronicles for Care Pathway Analysis

Participants: Yann Dauxais, Thomas Guyet, David Gross-Amblard [Druid], André Happe [Brest University Hospital].

A care pathway is a sequence of events (drugs deliveries, hospitalisation, etc) extracted from medical databases (see section 4.3 for details). In some studies, each patient is labeled by a class (*e.g.* died or not died). This information can be taken into account for the discriminant analysis of care pathways. This year, our objective was to extract discriminant patterns from a dataset of care pathways that can discriminate patients on their labels. To this end we introduced the new task of discriminant chronicle mining. Conceptually, a chronicle is a graph whose vertices are events and edges represent quantitative time constraints between events. We also proposed DCM, an algorithm dedicated to discriminant chronicles mining. This algorithm is based on rule learning methods to extract the temporal constraints. Computational performances and discriminant power of extracted chronicles are evaluated on artificial and real data.

The paper describing this work has been accepted in the french national conference on data mining (EGC 2017) [4] and is nominated for the best paper award.

7.4. Identifying Genetic Variant Combinations using Skypatterns

Participants: Alexandre Termier, Hoang-Son Pham [Genscale], Dominique Lavenier [Genscale].

Identifying variant combination association with disease is a bioinformatics challenge. This problem can be solved by discriminative pattern mining that uses a statistical function to evaluate the significance of individual biological patterns. There is a wide range of such measures. However, selecting an appropriate measure as well as a suitable threshold in some specific practical situations is a difficult task. In this work, we propose to use the skypattern technique which enables using combinations of measures to evaluate the importance of variant combinations without having to select a given measure and a fixed threshold (Pareto frontier). Experiments on several real variant datasets demonstrates that the skypattern method effectively identifies the risk variant combinations related to diseases [13].

7.5. Steady Patterns

Participants: Alexandre Termier, Willy Ugarte [UGA Grenoble], Miguel Santana [STMicroelectronics].

Skypatterns are an elegant answer to the pattern explosion issue, when a set of measures can be provided. Skypatterns for all possible measure combinations can be explored thanks to recent work on the skypattern cube. However, this leads to too many skypatterns, where it is difficult to quickly identify which ones are more important. First, we introduce a new notion of pattern steadiness [14] which measures the conservation of the skypattern property across the skypattern cube, allowing to see which are the "most universal" skypatterns. Then, we extended this notion to partitions of the dataset, and show in our experiments that this both allows to discover especially stable skypatterns, and identify interesting differences between the partitions.

7.6. Dense Bag-of-Temporal-SIFT-Words for Time Series Classification

Participants: Adeline Bailly [IRISA/Obelix], Laetitia Chapel [IRISA/Obelix], Thomas Guyet, Simon Malinowski [LinkMedia], Romain Tavenard [IRISA/Obelix].

The SIFT framework has shown to be effective in the image classification context. In [15], we designed a Bag-of-Words approach based on an adaptation of this framework to time series classification. It relies on two steps: SIFT-based features are first extracted and quantized into words; histograms of occurrences of each word are then fed into a classifier. In this work, we investigate techniques to improve the performance of Bag-of-Temporal-SIFT-Words: dense extraction of keypoints and different normalizations of Bag-of-Words histograms. Extensive experiments show that our method significantly outperforms nearly all tested standalone baseline classifiers on UCR datasets.

7.7. Comparing Symbolic and Statistical Classifiers on Energy Consumption Data

Participant: Benjamin Négrevergne.

During his Inria Carnot postdoc, Benjamin Négrevergne aimed at testing various data mining and machine learning methods on energy consumption data from the Energiency startup. Two symbolic methods developed in Lacodam were evaluated: QTempIntMiner and discriminant chronicle mining. While QTempIntMiner was shown to be ill-adapted in this setting, discriminant chronicle mining gave promising results. These results were evaluated in collaboration with our industrial partner. We also shown the interest of other methods: Hidden Markov Models and Gaussian processes. An internal report has been written to relate the results.

7.8. Detecting Strategic Moves in HearthStone Matches

Participants: Boris Doux [M1 intern], Clément Gautrais, Benjamin Negrevergne.

In this work, we demonstrate how to extract strategic knowledge from gaming data collected among players of the popular video game HearthStone. Our methodology is as follows. First we train a series of classifiers to predict the outcome of the game during a match, then we demonstrate how to spot key strategic events by tracking sudden changes in the classifier prediction. This methodology is applied to a large collection of HeathStone matches that we have collected from top ranked European players. Expert analysis shows that the events identified with this approach are both important and easy to interpret with the corresponding data [12].

7.9. Towards Visualizing Hidden Structures

Participants: Rémy Dautriche [STMicroelectronics], Alexandre Termier, Renaud Blanch [UGA Grenoble], Miguel Santana [STMicroelectronics].

There is an increasing need to quickly understand the contents of log data. A wide range of patterns can be computed and provide valuable information: for example existence of repeated sequences of events or periodic behaviors. However pattern mining techniques often produce many patterns that have to be examined one by one, which is time consuming for experts. On the other hand, visualization techniques are easier to understand, but cannot provide the in-depth understanding provided by pattern mining approaches. Our contribution is to propose a novel visual analytics method that allows to immediately visualize hidden structures such as repeated sets/sequences and periodicity, allowing to quickly gain a deep understanding of the log [3].

7.10. Knowledge-based Sequence Mining with ASP

Participants: Thomas Guyet, René Quiniou, Torsten Schaub.

We have introduced a framework for knowledge-based sequence mining, based on Answer Set Programming (ASP) [10], [5]. We begin by modeling the basic task and refine it in the sequel in several ways. First, we show how easily condensed patterns can be extracted by modular extensions of the basic approach. Second, we illustrate how ASP's preference handling capacities can be exploited for mining patterns of interest. In doing so, we demonstrate the ease of incorporating knowledge into the ASP-based mining process. To assess the trade-off in effectiveness, we provide an empirical study comparing our approach with a related sequence mining mechanism.

7.11. Packing Graphs with ASP for Landscape Simulation

Participants: Thomas Guyet, Yves Moinard, Jacques Nicolas [Dyliss], René Quiniou.

This work [6] describes an application of Answer Set Programming (ASP) to crop allocation for generating realistic landscapes. The task is to optimally cover a bare landscape, represented by its plot graph, with spatial patterns describing local arrangements of crops. This problem belongs to the hard class of graph packing problems and is modeled in the framework of ASP. The approach provides a compact and elegant solution to the basic problem and at the same time allows extensions such as a flexible integration of expert knowledge. Particular attention is paid to the treatment of symmetries, especially due to sub-graph isomorphism issues. Experiments were conducted on a database of simulated and real landscapes. Currently, the approach can process graphs of medium size, a size that enables studies on real agricultural practices.

7.12. Care Pathway Analysis with ASP Sequence Mining

Participants: Ahmed Samet, Benjamin Négrevergne, Thomas Guyet.

This line of work aims at applying our ASP encoding for sequential pattern mining to care pathway analysis (see section 4.3 for applicative objectives). This year, we proposed an approach of meaningful rare sequential pattern mining based on the declarative programming paradigm of Answer Set Programming (ASP). The setting of rare sequential pattern mining is introduced. To cope with the huge amount of meaningless rare patterns, our ASP approach provides an easy manner to encode expert constraints on expected patterns. Encodings are presented and quantitatively compared to a procedural baseline. An application on care pathways analysis illustrates the qualitative interest of expert constraints encoding.

This work has been submitted to the PAKDD 2017 conference.

7.13. ASP and Diagnosis

Participants: Christine Largouët, Laurence Rozé.

A new approach for performing diagnosis with ASP has been explored. The system is described by automata and implemented in an ASP program whose task is to find trajectories compatible with observations. The experimentation is carried out on benchmarks already used for the diagnosis problem using SAT. These benchmarks consider different levels of difficulty and number of faults (from one to twenty) and three types of observations: timestamped observations, totally ordered observations and partially ordered observations. The results were good both for dated and for totally ordered sequences of observations, whereas work needs to be still improved for the partial ordered observation case.

7.14. Searching for Cost-Optimized Strategies. Application to Temporal Planning and Agricultural System

Participants: Christine Largouët, Marie-Odile Cordier.

We consider a system modeled as a set of interacting components evolving along time according to explicit timing constraints. The decision making problem consists in selecting and organizing actions in order to reach a goal state in a limited time and in an optimal manner, assuming actions have a cost. We propose to reformulate the planning problem in terms of model-checking and controller synthesis such that the state to reach is expressed using a temporal logic. We have chosen to represent each agent using the formalism of Priced Timed Game Automata (PTGA) and a set of knowledge. PTGA is an extension of Timed Automata that allows the representation of cost on actions and the definition of a goal (to reach or to avoid). A first paper describes two algorithms designed to address the planning problem on a network of agents and proposes a practical implementation using model-checking tools that shows promising results on an agricultural application: a grassland based dairy production system [9]. Another paper describes the expressivity of this approach on the classical Transport Domain which is extended in order to include timing constraints, cost values and uncontrollable actions. This work has been implemented and performances evaluated on benchmarks [8].

7.15. Integrating Socio-Economic Drivers in an Explicit-Time, Qualitative Fisheries Model

Participant: Christine Largouët.

EcoMata is an explicit-time, qualitative modelling tool for assessing the ecosystem impacts of fishing and evaluating options for fishery management. The model is being developed further by integrating simple socio-economic drivers in the fishery system. Specifically, we have introduced a new module of automata that describes the profits associated to a specific fishing intensity and specific timing. This new module allows the evaluation of management strategies that are economically viable. The approach is illustrated on a coral-reef fishery in the Pacific that has been the focus of previous modelling work. [7].

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Contracts with Industry

8.1.1. SocTrace: analysis of SOC traces

Participant: Alexandre Termier.

SoCTrace is a FUI project led by STMicroelectronics, with the companies ProbaYes and Magilem, University Grenoble Alps and Inria Rhône-Alpes. Its goal is to provide an integrated environment for storing and analyzing execution traces. In this project, we are working on data mining techniques for analyzing the traces, and on the use of ontologies to enable querying traces with a higher level of abstraction.

8.1.2. ITRAMI: Interactive Trace Mining

Participant: Alexandre Termier.

ITRAMI is a Nano2017 project. Such projects are designed to support joint research efforts between STMicrolectronics and academic partners in the domain of embedded systems. Alexandre Termier is the PI of this project whose goal is to design novel data mining methods for interactive analysis of execution traces. Such methods aim at considerably reducing the time that STMicroelectronics developers spend at understanding, debugging and profiling applications running on STMicrolectronics chips. The project work is done at University Grenoble Alps, in collaboration with Lacodam researchers. Two contractual staff are working on the project in Grenoble: Willy Ugarte as a postdoc, and Soumaya Ben Alouane as an engineer.

8.2. Bilateral Grants with Industry

Maël Guillemé has obtained a CIFRE PhD grant with the Energiency startup, supervised by V. Masson and L. Rozé. The goal of Maël Guillemé's thesis is to propose new approaches for improving industrial energy performance and aims at integrating both numerical and symbolic attributes. A master 2 internship explored in 2016 a first approach based on an algorithm proposed by Shokoohi and al, but with several improvements: avoid data normalisation, detect patterns as fast as possible, enhance functions like distance and score.

Another CIFRE PhD has started, this time with the Amossys company, specialized in cyber-security. This is the PhD of Alban Siffer, located in the EMSec team of IRISA and co-supervised between EMSec (P.A. Fouque) and Lacodam (A. Termier, C. Largouët). The goal of this PhD is to propose new methods for intrusion detection in networks. The originality is to only consider IP flow as input (metadata of packets and not packet contents), requiring to detect intrusion via unusual traffic patterns.

9. Partnerships and Cooperations

9.1. Regional Initiatives

9.1.1. SePaDec: Declarative approaches for Sequential Pattern mining

Participants: Benjamin Negrevergne, Thomas Guyet, Ahmed Samet, Alexandre Termier.

The SEPADEC project is funded by the Region Bretagne. It aims at exploring the application of declarative pattern mining (more especially ASP) in the field of care pathway analysis. The first objective was to model knowledge from the data to enrich the raw data with medical expert knowledge and to develop a toolbox that smoothly integrates both expert knowledge and declarative pattern mining.

9.2. National Initiatives

9.2.1. ANR

9.2.1.1. #DigitAg: Digital agriculture

Participants: Alexandre Termier, Véronique Masson, Christine Largouët, Anne-Isabelle Graux.

#DigitAg is a “Convergence Institute” dedicated to the increasing importance of digital techniques in agriculture. Its goal is twofold: first, make innovative research on the use of digital techniques in agriculture in order to improve competitiveness, preserve environment, and offer correct living conditions to farmers. Second, through education prepare future farmers and agricultural policy makers to successfully exploit such technology.

While #DigitAg is based on Montpellier, Rennes is a satellite of the institute focused on cattle farming. Lacodam is involved in the “data mining” challenge of the institute, that A. Termier co-leads. He is also the representative of Inria in the steering comittee of the institute.

The interest for the team is to design novel methods to analyze and represent agricultural data, which are challenging because they are both heterogeneous and multi-scale (both spatial and temporal).

9.2.2. National Platforms

9.2.2.1. PEPS: Pharmaco-epidemiology for Health Products

Participants: Yann Dauxais, Thomas Guyet, Véronique Masson, René Quiniou, Alexandre Termier.

The PEPS project (Pharmaco-epidemiology des Produits de Santé) is funded by ANSM (national agency for health security). The project leader is E. Oger from the clinical investigation center CIC-1414 INSERM/CHU Rennes. The other partners located in Rennes are the Institute of Research and Technology (IRT) B<>Com, EHESP and the LTSI. The project started in january 2015 and is funded for 4 years.

The PEPS project has two parts: the clinical studies and a research program dedicated to the development of innovative tools for pharmaco-epidemiological studies with medico-administrative databases.

Our contribution to this project will be to propose pattern mining algorithm and reasoning techniques to analyse the typical care pathways of specific groups of insured patients.

9.3. International Initiatives

9.3.1. Inria International Partners

9.3.1.1. Informal International Partners

9.3.1.1.1. University of Calgary: Monitoring cattle in big herds with multiple sensors

Participant: René Quiniou.

The state of Alberta produces a significant part of the beef meat in Canada. Big farms feeds up around 40.000 bull calves in feedlots grouping 200-300 animals. Diseases such as Bovine Respiratory Diseases (BRD) are frequent and may propagate quickly in such conditions. So, it is important to detect as soon as possible when an animal is sick. We are collaborating with the Department of Production Animal Health, University of Calgary for designing monitoring systems able to generate early alarms when an animal is sick. Precisely, we are studying the properties of new sensors and their aptitude to provide relevant data for BRD detectors.

9.3.1.1.2. University of Potsdam: preferences in mining with ASP

Participant: Thomas Guyet.

The research group “knowledge processing and information systems" of the University of Potsdam, so called Potascco group, develops a collection of tools and programs for Answer Set Programming such as the clingo solver or the ASPRIN system, developed by J. Romero to handle preferences on ASP models. They have strong expertise in problem encoding with ASP. In addition to T. Schaub Inria position, we initiate some collaborations with other members of the Potascco group in order to strengthen our relationships.

9.4. International Research Visitors

9.4.1. Research Stays Abroad

Thomas Guyet spent a month (may 2016) in the team leaded by Prof. Torsten Schaub in the university of Potsdam.

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events Organisation

10.1.1.1. Member of the Organizing Committees

- Local chair of the Mini-symposium on instant data mining, interactive data mining, preference-based pattern mining, Rennes, October 26-28, 2016 (A. Termier)
- Organization chair (T. Guyet) and program committee members (T. Guyet, R. Quiniou) of GAST workshop at EGC 2017 and at EGC 2016.

10.1.2. Scientific Events Selection

10.1.2.1. Member of the Conference Program Committees

EGC 2017 (T. Guyet, R. Quiniou, A. Termier).

ICDM 2016 (A. Termier)

KDD 2016 (A. Termier)

IJCAI 2016 (T. Guyet, B. Negrevergne)

ECAI 2016 (T. Guyet)

KR 2016 (T. Guyet)

RFIA/CNIA 2016 (T. Guyet)

APIA 2016 (C. Largouët)

10.1.2.2. Reviewer

R. Quiniou: KDD 2016, ECAI 2016, IJCAI 2016, ICDM 2016, CNIA 2016, AKDM7

T. Guyet: KDD 2016, ECAI 2016, IJCAI 2016, ICDM 2016, RFIA/CNIA 2016, KR 2016, AALTD16, SimBig 2016

A. Termier: IJCAI 2016

Y. Dauxais, C. Gautrais: IJCAI 2016, ICDM 2016, KDD 2016, EGC 2016

10.1.3. Journal

10.1.3.1. Member of the Editorial Boards

T. Guyet: RIA (Revue d'Intelligence Artificielle)

10.1.3.2. Reviewer - Reviewing Activities

A. Termier: Data Mining and Knowledge Discovery, Knowledge and Information Systems, Discrete Applied Mathematics

T. Guyet: Journal of Biomedical Informatics, AI Communications, ACM Computing Surveys, TKDE

C. Gautrais: Data Mining and Knowledge Discovery

10.1.4. Invited Talks

A. Termier gave an invited talk at the SMiLe 2016 workshop.

C. Largouët gave a invited talk at the UMR-AMURE Seminar on the 15th september 2016,

10.1.5. Leadership within the Scientific Community

A. Termier is the representative of Inria for the #DigitAg Convergence Institue

T. Guyet is member of the AFIA board (since october 2011). AFIA is the french chapter of ECCAI, the European Association for Artificial Intelligence.

10.1.6. Scientific Expertise

Evaluation of a project proposal for the ANR: R. Quiniou

Evaluation of a project proposal for the Medicen competitive cluster: T. Guyet

10.1.7. Research Administration

Member of INRA CEI (Engineers Evaluation Committee): T. Guyet

Member of the scientific board of department EA of INRA: A. Termier

Member of the scientific board of Agrocampus Ouest - COREGE: C. Largouët.

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

Many members of the project-team Lacodam are also faculty members and are actively involved in computer science teaching programs in ISTIC, INSA and Agrocampus-Ouest. Besides these usual teachings Lacodam is involved in the following programs:

Master 2 Module DSS: Apprentissage sur des données séquentielles symboliques, 10 h, M2, Istic University of Rennes 1 (R. Quiniou)

Master 2 Géoinformation, Agrocampus Ouest Rennes (L. Bonneau, T. Guyet)

Master 2 Artificial Intelligence, Agrocampus Ouest Rennes (Louis Bonneau, C. Largouët)

Master 1 Scientific Programming, Data Management, Python Programming, Agrocampus Ouest Rennes (C. Largouët)

10.2.2. Supervision

PhD in progress: Maël Guillemé, “New data mining approaches for improving energy consumption in factory"”, october 3rd 2016, Alexandre Termier, Véronique, Masson and Laurence Rozé

PhD in progress: Clément Gautrais, “Mining massive data from client purchases”, october 1st 2015, Alexandre Termier, Peggy Cellier, Thomas Guyet and René Quiniou

PhD in progress: Yann Dauxais, "“Query-language for care-pathway mining and analysis”, february 1st 2015, David Gross-Amblard, Thomas Guyet, André Happe

PhD in progress: Alban Siffer, “DataMining approaches for cyber attack detection”, mars 2016, Pierre-Alain Fouque, Alexandre Termier, Christine Largouët

10.2.3. Juries

Committee member of Rémy Dautriche Phd defense (Université de Grenoble Alpes): A. Termier

Reviewer of Olivier Cavadenti Phd (INSA Lyon): A. Termier

Committee member of Samir Loudni HDR (Université de Caen): A. Termier

Thesis advisory committee member of Jean Coquet (Univ. Rennes 1): A. Termier

Thesis advisory committee member of Benoit Bellot (INRA/IGEPP): T. Guyet

Thesis advisory committee member of Zhi Cheng (UNC/PPME): T. Guyet

10.3. Popularization

M.-O. Cordier is editorial board member of Interstices webzine.

11. Bibliography

Major publications by the team in recent years

[1] T. GUYET, R. QUINIOU.*Extracting temporal patterns from interval-based sequences*, in "International Join Conference on Artificial Intelligence (IJCAI)", Barcelone, Spain, July 2011, https://hal.inria.fr/inria-00618444.

Publications of the year

Articles in International Peer-Reviewed Journal

[2] H. MIRISAEE, E. GAUSSIER, A. TERMIER.*Efficient local search for L_1 and L_2 binary matrix factorization*, in "Intelligent Data Analysis", 2016, vol. 20, p. 783 - 807 [*DOI :* 10.3233/IDA-160832], https://hal.archives-ouvertes.fr/hal-01405186.

International Conferences with Proceedings

[3] R. DAUTRICHE, A. TERMIER, R. BLANCH, M. SANTANA.*Towards Visualizing Hidden Structures*, in "International Conference on Data Mining (ICDM) / PhD Forum", Barcelone, Spain, 2016, https://hal.archives-ouvertes.fr/hal-01407664.

[4] Y. DAUXAIS, D. GROSS-AMBLARD, T. GUYET, A. HAPPE.*Extraction de chroniques discriminantes*, in "Extraction et Gestion des Connaissances (EGC)", Grenoble, France, January 2017, https://hal.inria.fr/hal-01413473.

[5] M. GEBSER, T. GUYET, R. QUINIOU, J. ROMERO, T. SCHAUB.*Knowledge-based Sequence Mining with ASP*, in "IJCAI 2016- 25th International joint conference on artificial intelligence", New-york, United States, AAAI, July 2016, 8, https://hal.inria.fr/hal-01327363.

[6] T. GUYET, Y. MOINARD, J. NICOLAS, R. QUINIOU.*Packing graphs with ASP for landscape simulation*, in "IJCAI 2016 - 25th International joint conference on artificial intelligence", New-york, United States, AAAI, July 2016, 8, https://hal.inria.fr/hal-01327368.

[7] C. LARGOUËT, Y.-M. BOZEC, Y. ZHAO.*Integrating socio-economic drivers in an explicit-time, qualitative fisheries model: EcoMata*, in "Understanding Marine Socio-ecological Systems including the human dimension in integrated ecosystem assessment", Brest, France, ICES, May 2016, https://hal.archives-ouvertes.fr/hal-01398963.

[8] C. LARGOUËT, O. KRICHEN, Y. ZHAO.*Temporal Planning with extended Timed Automata*, in "28th International Conference on Tools with Artificial Intelligence (ICTAI 2016)", SAN JOSE, United States, N. BOURBAKIS, A. ESPOSITO, A. MALI, M. ALAMANIOTIS (editors), IEEE, November 2016, https://hal.archives-ouvertes.fr/hal-01398695.

[9] C. LARGOUËT, Y. ZHAO, M.-O. CORDIER.*Searching for Cost-Optimized Strategies: An Agricultural Application*, in "2nd International Conference, ICDSST 2016", Plymouth, United Kingdom, Decision Support Systems VI - Addressing Sustainability and Societal Challenges Volume 250, Springer, May 2016, vol. Lecture Notes in Business Information Processing, n^{o} 250, p. 31 - 43, Best Paper Award [*DOI :* 10.1007/978-3-319-32877-5_3], https://hal.archives-ouvertes.fr/hal-01398660.

National Conferences with Proceeding

[10] T. Guyet, Y. Moinard, R. Quiniou, T. Schaub.*Fouille de motifs séquentiels avec ASP*, in "Extraction et Gestion de Connaissances (EGC)", Reims, France, Actes de la conférence Extraction et Gestion de Connaissances, RNTI, 2016, https://hal.inria.fr/hal-01239501.

Conferences without Proceedings

[11] F. Balusson, M.-A. Botrel, O. Dameron, Y. Dauxais, E. Drezen, A. Dupuy, T. Guyet, D. Gross-Amblard, A. Happe, N. Le Meur, B. Le Nautout, E. Leray, E. Nowak, C. Rault, E. Oger, E. Polard.*PEPS: a platform for supporting studies in pharmaco-epidemiology using medico-administrative databases*, in "International Congress on e-Health Research", Paris, France, October 2016, https://hal.inria.fr/hal-01380939.

[12] B. Doux, C. Gautrais, B. Negrevergne.*Detecting strategic moves in HearthStone matches*, in "Machine Learning and Data Mining for Sports Analytics Workshop of ECML/PKDD", Riva del Garda, Italy, September 2016, https://hal.archives-ouvertes.fr/hal-01412432.

[13] H.-S. Pham, D. Lavenier, A. Termier.*Identifying Genetic Variant Combinations using Skypatterns*, in "7th International Workshop on Biological Knowledge Discovery and Data Mining (Workshop BIOKDD '16)", Porto, Portugal, DEXA, September 2016 [*DOI :* 10.1109/DEXA.2016.13], https://hal.inria.fr/hal-01385614.

[14] W. Ugarte, A. Termier, M. Santana.*Steady Patterns*, in "Data Science and Big Data Analytics workshop of International Conference on Data Mining", Barcelone, Spain, 2016, https://hal.archives-ouvertes.fr/hal-01408397.

Scientific Books (or Scientific Book chapters)

[15] A. Bailly, S. Malinowski, R. Tavenard, L. Chapel, T. Guyet.*Dense Bag-of-Temporal-SIFT-Words for Time Series Classification*, in "Advanced Analysis and Learning on Temporal Data", Springer, 2016, https://hal.archives-ouvertes.fr/hal-01252726.

Other Publications

[16] C. Gautrais, P. Cellier, T. Guyet, R. Quiniou, A. Termier.*Understanding Customer Attrition at an Individual Level: a New Model in Grocery Retail Context*, March 2016, International Conference on Extending Database Technology (EDBT), Poster [*DOI :* 10.5441/002/EDBT.2016.87], https://hal.archives-ouvertes.fr/hal-01405172.

References in notes

[17] S. Colas, C. Collin, P. Piriou, M. Zureik.*Association between total hip replacement characteristics and 3-year prosthetic survivorship: A population-based study*, in "JAMA Surgery", 2015, vol. 150, n° 10, p. 979–988.

[18] G. Moulis, M. Lapeyre-Mestre, A. Palmaro, G. Pugnet, J.-L. Montastruc, L. Sailler.*French health insurance databases: What interest for medical research?*, in "La Revue de Médecine Interne", 2015, vol. 36, n° 6, p. 411 - 417.

[19] E. NOWAK, A. HAPPE, J. BOUGET, F. PAILLARD, C. VIGNEAU, P.-Y. SCARABIN, E. OGER.*Safety of Fixed Dose of Antihypertensive Drug Combinations Compared to (Single Pill) Free-Combinations: A Nested Matched Case–Control Analysis*, in "Medicine", 2015, vol. 94, n^o 49, e2229.

[20] E. POLARD, E. NOWAK, A. HAPPE, A. BIRABEN, E. OGER.*Brand name to generic substitution of antiepileptic drugs does not lead to seizure-related hospitalization: a population-based case-crossover study*, in "Pharmacoepidemiology and drug safety", 2015, vol. 24, n^o 11, p. 1161–1169.

Project-Team LAGADIC

Visual servoing in robotics, computer vision, and augmented reality

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

IN PARTNERSHIP WITH:
CNRS

Institut national des sciences appliquées de Rennes

Université Rennes 1

RESEARCH CENTERS
Rennes - Bretagne-Atlantique
Sophia Antipolis - Méditerranée

THEME
Robotics and Smart environments

Table of contents

Project-Team LAGADIC

Creation of the Project-Team: 2004 December 06

Keywords:

Computer Science and Digital Science:

5.4.4. - 3D and spatio-temporal reconstruction
5.4.5. - Object tracking and motion analysis
5.4.6. - Object localization
5.4.7. - Visual servoing
5.6. - Virtual reality, augmented reality
5.10.2. - Perception
5.10.4. - Robot control
5.10.5. - Robot interaction (with the environment, humans, other robots)
5.10.6. - Swarm robotics

Other Research Topics and Application Domains:

2.4.3. - Surgery
2.5. - Handicap and personal assistances
5.1. - Factory of the future
5.6. - Robotic systems
7.2.1. - Smart vehicles
8.4. - Security and personal assistance

1. Members

Research Scientists

François Chaumette [Team leader, Inria, Senior Researcher, HDR]
Alexandre Krupa [Inria, Researcher, HDR]
Claudio Pacchierotti [CNRS, Researcher, from Dec 2016]
Julien Pettré [Inria, Researcher, from April 2016, HDR]
Patrick Rives [Inria, Senior Researcher, HDR]
Paolo Robuffo Giordano [CNRS, Senior Researcher, HDR]
Paolo Salaris [Inria, Researcher]

Faculty Members

Marie Babel [INSA Rennes, Associate Professor, HDR]
Vincent Drevelle [Univ. Rennes I, Associate Professor]
Eric Marchand [Univ. Rennes I, Professor, HDR]

Technical Staff

Fabien Spindler [Inria, Senior engineer]
Thomas Bellavoir [CNRS, granted by ANR SenseFly project]
Giovanni Claudio [Inria, granted by BPI Romeo 2 project and H2020 Comanoid project]
Pierre-Marie Kerzerho [Inria, granted by ANR VisioLand, from Nov 2016]
Andrea Peruffo [CNRS, granted by H2020 RoMans, until Apr 2016]
Marc Pouliquen [Inria, granted by ADT UsTk, from Sep 2016]
Souriya Trinh [Inria, granted by H2020 Comanoid project]
Aurélien Yol [Inria, granted by ANR VisioLand project]

PhD Students

Firas Abi Farraj [CNRS, granted by H2020 Romans project]
Quentin Bateux [Univ. Rennes I, granted by MESR]
Aline Baudry [INSA Rennes, granted by MESR and Brittany Council, from Oct 2016]
Suman Raj Bista [Inria, granted by BPI Romeo 2 and Brittany Council]
Julien Bruneau [Univ. Rennes I, granted by MESR]
Nicolas Cazy [Inria, granted by BPI Romeo 2 project]
Pierre Chatelain [Univ. Rennes 1, granted by ENS Cachan]
Jason Chevrie [Univ. Rennes I, granted by ENS Cachan]
Le Cui [Univ. Rennes I, granted by ANR Nanorobust, until Jan 2016]
Quentin Delamare [Univ. Rennes I, granted by ENS Rennes, from Sep 2016]
Louise Devigne [Pôle Saint Hélier, granted by Cifre]
Lesley-Ann Duflot [Inria, granted by Femto-ST and Brittany Council]
Hadrien Gurnel [Inst. de Recherche Technologique b<>com, from Oct 2016]
Dayana Hassan [Inria, granted by Cifre, from Nov 2016]
Salma Jiddi [Technicolor, granted by Cifre]
Vishnu Karakkat Narayanan [Inria, granted by AEN Pal]
Ide Flore Kenmogne Fokam [Inria, granted in part by Brittany Council]
Axel Lopez Gandia [Univ. Rennes I, granted by MESR, from Oct 2016]
Aly Magassouba [Univ. Rennes I, granted by MESR]
Renato José Martins [CNPq,]
Noël Mériaux [Inria, granted by ANR VisioLand]
Pedro Alfonso Patlan Rosales [Conacyt]
Bryan Penin [Inria, granted in part by DGA]
Lucas Royer [Inst. de Recherche Technologique b<>com]
Fabrizio Schiano [Inria]
Muhammad Usman [CNRS, granted by ANR SenseFly]

Post-Doctoral Fellows

Don Joven Agravante [Inria, granted by H2020 Comanoid project]
Nicolò Pedemonte [CNRS, granted by H2020 Romans project, until Sep 2016]
Riccardo Spica [CNRS, granted by ANR SenseFly]
David Wolinski [Inria, until Nov 2016, granted by ANR Percolation]
Panayiotis Charalambous [Inria, granted by ANR Percolation, until May 2016]
Eduardo Fernandez Moral [Inria, granted by ANR Platinum]

Administrative Assistants

Hélène de La Ruée [Inria, in common with I4S group]
Christine Riehl [Inria, in common with Hephaistos group and Focus group at Univ. of Bologna]

2. Overall Objectives

2.1. Overall Objectives

Historically, research activities of the Lagadic team are concerned with visual servoing, visual tracking, and active vision. Visual servoing consists in using the information provided by a vision sensor to control the movements of a dynamic system. This research topic is at the intersection of the fields of robotics, automatic control, and computer vision. These fields are the subject of profitable research since many years and are particularly interesting by their very broad scientific and application spectrum. Within this spectrum, we focus on the interaction between visual perception and action. This topic is significant because it provides an alternative to the traditional Perception-Decision-Action cycle. It is indeed possible to link the perception and action aspects more closely, by directly integrating the measurements provided by a vision sensor in closed loop control laws. Our objective is thus to design strategies of coupling perception and action from images for applications in robotics, computer vision, virtual reality and augmented reality.

This objective is significant, first of all because of the variety and the great number of potential applications to which our work can lead (see Section 4.1). Secondly, it is also significant to be able to raise the scientific aspects associated with these problems, namely modeling of visual features representing the interaction between action and perception in an optimal way, taking into account of complex environments and the specification of high level tasks. We also work to treat new problems provided by imagery systems such as those resulting from an omnidirectional vision sensor or echographic probes. We are finally interested in revisiting traditional problems in computer vision (3D localization) through the visual servoing approach.

Thanks to the arrival of Patrick Rives and his students in the group in April 2012, which makes Lagadic now localized both in Rennes and Sophia Antipolis, the group now also focuses on building consistent representations of the environment that can be used to trigger and execute the robot actions. In its broadest sense, perception requires detecting, recognizing, and localizing elements of the environment, given the limited sensing and computational resources available on the embedded system. Perception is a fundamental issue for both the implementation of reactive behaviors, as is traditionaly studied in the group, and the construction of the representations that are used at the task level. Simultaneous Localization and Mapping (Slam) is thus now one of our research areas.

Among the sensory modalities, computer vision, range finder and odometry are of particular importance and interest for mobile robots due to their availability and extended range of applicability, while ultrasound images and force measurements are both required for our medical robotics applications. The fusion of complementary information provided by different sensors is thus also a central issue for modeling the environment, robot localization, control, and navigation.

Much of the processing must be performed in real time, with a good degree of robustness so as to accommodate with the large variability of the physical world. Computational efficiency and well-posedness of the methods developed are thus constant preoccupations of the group.

3. Research Program

3.1. Visual servoing

Basically, visual servoing techniques consist in using the data provided by one or several cameras in order to control the motions of a dynamic system [1]. Such systems are usually robot arms, or mobile robots, but can also be virtual robots, or even a virtual camera. A large variety of positioning tasks, or mobile target tracking, can be implemented by controlling from one to all the degrees of freedom of the system. Whatever the sensor configuration, which can vary from one on-board camera on the robot end-effector to several free-standing cameras, a set of visual features has to be selected at best from the image measurements available, allowing to control the desired degrees of freedom. A control law has also to be designed so that these visual features $\mathbf{s}(t)$ reach a desired value $\mathbf{s}^*$, defining a correct realization of the task. A desired planned trajectory $\mathbf{s}^*(t)$ can also be tracked. The control principle is thus to regulate the error vector $\mathbf{s}(t) - \mathbf{s}^*(t)$ to zero. With a vision sensor providing 2D measurements, potential visual features are numerous, since 2D data (coordinates of feature points in the image, moments, ...) as well as 3D data provided by a localization algorithm exploiting the extracted 2D features can be considered. It is also possible to combine 2D and 3D visual features to take the advantages of each approach while avoiding their respective drawbacks.

More precisely, a set $\mathbf{s}$ of k visual features can be taken into account in a visual servoing scheme if it can be written:

$$\mathbf{s} = \mathbf{s}(\mathbf{x}(\mathbf{p}(t)), \mathbf{a}) \tag{54}$$

where $\mathbf{p}(t)$ describes the pose at the instant t between the camera frame and the target frame, $\mathbf{x}$ the image measurements, and $\mathbf{a}$ a set of parameters encoding a potential additional knowledge, if available (such as for instance a coarse approximation of the camera calibration parameters, or the 3D model of the target in some cases).

The time variation of $\mathbf{s}$ can be linked to the relative instantaneous velocity $\mathbf{v}$ between the camera and the scene:

$$\dot{\mathbf{s}} = \frac{\partial \mathbf{s}}{\partial \mathbf{p}}\, \dot{\mathbf{p}} = \mathbf{L_s}\, \mathbf{v} \tag{55}$$

where $\mathbf{L_s}$ is the interaction matrix related to $\mathbf{s}$. This interaction matrix plays an essential role. Indeed, if we consider for instance an eye-in-hand system and the camera velocity as input of the robot controller, we obtain when the control law is designed to try to obtain an exponential decoupled decrease of the error:

$$\mathbf{v}_c = -\lambda \widehat{\mathbf{L_s}}^{+} (\mathbf{s} - \mathbf{s}^*) - \widehat{\mathbf{L_s}}^{+} \widehat{\frac{\partial \mathbf{s}}{\partial t}} \tag{56}$$

where λ is a proportional gain that has to be tuned to minimize the time-to-convergence, $\widehat{\mathbf{L_s}}^{+}$ is the pseudo-inverse of a model or an approximation of the interaction matrix, and $\widehat{\frac{\partial \mathbf{s}}{\partial t}}$ an estimation of the features velocity due to a possible own object motion.

From the selected visual features and the corresponding interaction matrix, the behavior of the system will have particular properties as for stability, robustness with respect to noise or to calibration errors, robot 3D trajectory, etc. Usually, the interaction matrix is composed of highly non linear terms and does not present any decoupling properties. This is generally the case when $\mathbf{s}$ is directly chosen as $\mathbf{x}$. In some cases, it may lead to inadequate robot trajectories or even motions impossible to realize, local minimum, tasks singularities, etc. It is thus extremely important to design adequate visual features for each robot task or application, the ideal case (very difficult to obtain) being when the corresponding interaction matrix is constant, leading to a simple linear control system. To conclude in a few words, **visual servoing is basically a non linear control problem. Our Holy Grail quest is to transform it into a linear control problem.**

Furthermore, embedding visual servoing in the task function approach allows solving efficiently the redundancy problems that appear when the visual task does not constrain all the degrees of freedom of the system. It is then possible to realize simultaneously the visual task and secondary tasks such as visual inspection, or joint limits or singularities avoidance. This formalism can also be used for tasks sequencing purposes in order to deal with high level complex applications.

3.2. Visual tracking

Elaboration of object tracking algorithms in image sequences is an important issue for researches and applications related to visual servoing and more generally for robot vision. A robust extraction and real time spatio-temporal tracking process of visual cues is indeed one of the keys to success of a visual servoing task. If fiducial markers may still be useful to validate theoretical aspects in modeling and control, natural scenes with non-cooperative objects and subject to various illumination conditions have to be considered for addressing large scale realistic applications.

Most of the available tracking methods can be divided into two main classes: feature-based and model-based. The former approach focuses on tracking 2D features such as geometrical primitives (points, segments, circles,...), object contours, regions of interest, etc. The latter explicitly uses a model of the tracked objects. This can be either a 3D model or a 2D template of the object. This second class of methods usually provides a more robust solution. Indeed, the main advantage of the model-based methods is that the knowledge about the scene allows improving tracking robustness and performance, by being able to predict hidden movements of the object, detect partial occlusions and acts to reduce the effects of outliers. The challenge is to build algorithms that are fast and robust enough to meet our application requirements. Therefore, even if we still consider 2D feature tracking in some cases, our researches mainly focus on real-time 3D model-based tracking, since these approaches are very accurate, robust, and well adapted to any class of visual servoing schemes. Furthermore, they also meet the requirements of other classes of application, such as augmented reality.

3.3. Slam

Most of the applications involving mobile robotic systems (ground vehicles, aerial robots, automated submarines,...) require a reliable localization of the robot in its environment. A challenging problem is when neither the robot localization nor the map is known. Localization and mapping must then be considered concurrently. This problem is known as Simultaneous Localization And Mapping (Slam). In this case, the robot moves from an unknown location in an unknown environment and proceeds to incrementally build up a navigation map of the environment, while simultaneously using this map to update its estimated position.

Nevertheless, solving the Slam problem is not sufficient for guaranteeing an autonomous and safe navigation. The choice of the representation of the map is, of course, essential. The representation has to support the different levels of the navigation process: motion planning, motion execution and collision avoidance and, at the global level, the definition of an optimal strategy of displacement. The original formulation of the Slam problem is purely metric (since it basically consists in estimating the Cartesian situations of the robot and a set of landmarks), and it does not involve complex representations of the environment. However, it is now well recognized that **several complementary representations are needed to perform exploration, navigation, mapping, and control tasks successfully. We propose to use composite models of the environment that mix topological, metric, and grid-based representations.** Each type of representation is well adapted to a particular aspect of autonomous navigation [7]: the metric model allows one to locate the robot precisely and plan Cartesian paths, the topological model captures the accessibility of different sites in the environment and allows a coarse localization, and finally the grid representation is useful to characterize the free space and design potential functions used for reactive obstacle avoidance. However, ensuring the consistency of these various representations during the robot exploration, and merging observations acquired from different viewpoints by several cooperative robots, are difficult problems. This is particularly true when different sensing modalities are involved. New studies to derive efficient algorithms for manipulating the hybrid representations (merging, updating, filtering...) while preserving their consistency are needed.

3.4. Scene modeling and understanding

Long-term mapping has received an increasing amount of attention during last years, largely motivated by the growing need to integrate robots into the real world wherein dynamic objects constantly change the appearance of the scene. A mobile robot evolving in such a dynamic world should not only be able to build a map of the observed environment at a specific moment, but also to maintain this map consistent over a long period of time. It has to deal with dynamic changes that can cause the navigation process to fail. However updating the map is particularly challenging in large-scale environments. To identify changes, robots have to keep a memory of the previous states of the environment and the more dynamic it is, the higher will be the number of states to manage and the more computationally intensive will be the updating process. Mapping large-scale dynamic environments is then particularly difficult as the map size can be arbitrary large. Additionally, mapping many times the whole environment is not always possible or convenient and it is useful to take advantages of methods using only a small number of observations.

A recent trend in robotic mapping is to augment low-level maps with semantic interpretation of their content, which allows to improve the robot's environmental awareness through the use of high-level concepts. In mobile robot navigation, the so-called semantic maps have already been used to improve path planning methods, mainly by providing the robot with the ability to deal with human-understandable targets.

4. Application Domains

4.1. Application Domains

The natural applications of our research are obviously in robotics. In fact, researches undertaken in the Lagadic group can apply to all the fields of robotics implying a vision sensor. They are indeed conceived to be independent of the system considered (and the robot and the vision sensor can even be virtual for some applications).

Currently, we are mostly interested in using visual servoing for aerial and space application, micromanipulation, autonomous vehicle navigation in large urban environments or for disabled or elderly people.

We also address the field of medical robotics. The applications we consider turn around new functionalities of assistance to the clinician during a medical examination: visual servoing on echographic images, needle insertion, compensation of organ motion, etc.

Robotics is not the only possible application field to our researches. In the past, we were interested in applying visual servoing in computer animation, either for controlling the motions of virtual humanoids according to their pseudo-perception, or for controlling the point of view of visual restitution of an animation. In both cases, potential applications are in the field of virtual reality, for example for the design of video games, or virtual cinematography.

Applications also exist in computer vision and augmented reality. It is then a question of carrying out a virtual visual servoing for the 3D localization of a tool with respect to the vision sensor, or for the estimation of its 3D motion. This field of application is very promising, because it is in full rise for the realization of special effects in the multi-media field or for the design and the inspection of objects manufactured in the industrial world.

5. Highlights of the Year

5.1. Highlights of the Year

- Eric Marchand and Fabien Spindler co-authored with Prof. Hideaki Uchiyama (Kyushu Univ., Japan) a survey on pose estimation for augmented reality published in IEEE Trans. on Visualization and Computer Graphics [33].
- The second edition of the Springer Handbook of Robotics has been released this year. It contains an extended version of the chapter on visual servoing co-authored by François Chaumette, Prof. Seth Hutchinson (UIUC, Illinois) and Prof. Peter Corke (QUT, Brisbane, Australia) [77].

5.1.1. Awards

- The ANR project ENTRACTE, of which Julien Pettré is partner, has received the “ANR Grand Prix du Numérique 2016”. The project is about anthropomorphic action planning and understanding: http://www.agence-nationale-recherche.fr/?Project=ANR-13-CORD-0002 (see also Section 9.2.3).
- Paper [71] has been selected has one of the five finalists for the ICARCV’2016 Best Paper Award.
- Lagadic is a member of the five finalist teams for the KUKA Innovation Award (https://www.kuka.com/en-de/press/events/kuka-innovation-award), together with the RIS group at LAAS (coordinator), the University of Siena, Italy, and the Seoul National University, South Korea. The goal is to address search and rescue operations in regions which are difficult to access or dangerous following disasters. For this, the team will explore the collaboration between a quadrotor UAV and a KUKA lightweight arm for cooperative transportation and manipulation of rigid objects (e.g., long bards), with a final peg-in-hole task to be demonstrated live at the Hannover fair during spring 2017.

6. New Software and Platforms

6.1. DESlam

Dense Egocentric SLAM
KEYWORDS: Deph Perception - Robotics - Localization
FUNCTIONAL DESCRIPTION

This software proposes a full and self content solution to the dense Slam problem. Based on a generic RGB-D representation valid for various types of sensors (stereovision, multi-cameras, RGB-D sensors...), it provides a 3D textured representation of complex large indoor and outdoor environments and it allows localizing in real time (45Hz) a robot or a person carrying out a mobile camera.

- Participants: Maxime Meilland, Andrew Ian Comport and Patrick Rives
- Contact: Patrick Rives
- URL: http://team.inria.fr/lagadic

6.2. HandiViz

KEYWORDS: Health - Persons attendant - Handicap
FUNCTIONAL DESCRIPTION

The HandiViz software proposes a semi-autonomous navigation framework of a wheelchair relying on visual servoing.

It has been registered to the APP ("Agence de Protection des Programmes") as an INSA software (IDDN.FR.001.440021.000.S.P.2013.000.10000) and is under GPL license.

- Participants: Francois Pasteau and Marie Babel
- Contact: Marie Babel
- URL: https://team.inria.fr/lagadic/

6.3. Perception360

Robot vision and 3D mapping with omnidirectional RGB-D sensors.
KEYWORDS: Depth Perception - 3D rendering - Computer vision - Robotics - Image registration - Sensors - Realistic rendering - 3D reconstruction - Localization
FUNCTIONAL DESCRIPTION

This software is a collection of libraries and applications for robot vision and 3D mapping with omnidirectional RGB-D sensors or standard perspective cameras. It provides the functionalities to do image acquisition, semantic annotation, dense registration, localization and 3D mapping. The omnidirectional RGB-D sensors used within this software have been developed at Inria Sophia Antipolis.

- Participants: Eduardo Fernandez Moral, Renato José Martins and Patrick Rives
- Contact: Patrick Rives
- URL: https://team.inria.fr/lagadic

6.4. Sinatrack

KEYWORDS: Computer vision - Robotics
FUNCTIONAL DESCRIPTION

Sinatrack is a tracking software that performs the 3D localization (translation and rotation) of an object with respect to a monocular camera. It allows considering objects with complex shape. The underlying approach is a model-based tracking technique. It has been developed for satellite localization and on-orbit service applications but is also suitable for augmented reality purpose.

- Participants: Antoine Petit, Eric Marchand and Francois Chaumette
- Contact: Eric Marchand
- URL: http://team.inria.fr/lagadic

6.5. UsTk

Ultrasound Toolkit

KEYWORDS: Echographic imagery - Image reconstruction - Active contours - Medical robotics

FUNCTIONAL DESCRIPTION

UsTk, standing for Ultrasound Toolkit, is a cross-platform library for two- and three-dimensional ultrasound image processing and visual servoing based on ultrasound images. Written in C++, UsTk provides tools for ultrasound image acquisition, processing, and display, as well as control of ultrasound probe motion by ultrasound visual servoing. This year we started the development of a new version. The objective is first to consolidate existing developments, to improve the quality of the software, to add new state-of-the-art algorithms, and then to disseminate them within the community as an open-source software.

- Participants: Marc Pouliquen, Alexandre Krupa, Pierre Chatelain and Fabien Spindler
- Contact: Alexandre Krupa
- URL: https://team.inria.fr/lagadic/

6.6. ViSP

KEYWORDS: Computer vision - Robotics - Augmented reality - Visual servoing

SCIENTIFIC DESCRIPTION

Since 2005, we have been developing and releasing ViSP [5], an open source library available from http://visp.inria.fr. ViSP standing for Visual Servoing Platform allows prototyping and developing applications using visual tracking and visual servoing techniques at the heart of the Lagadic research. ViSP was designed to be independent from the hardware, to be simple to use, expandable and cross-platform. ViSP allows to design vision-based tasks for eye-in-hand and eye-to-hand visual servoing that contains the most classical visual features that are used in practice. It involves a large set of elementary positioning tasks with respect to various visual features (points, segments, straight lines, circles, spheres, cylinders, image moments, pose...) that can be combined together, and image processing algorithms that allow tracking of visual cues (dots, segments, ellipses...) or 3D model-based tracking of known objects or template tracking. Simulation capabilities are also available.

FUNCTIONAL DESCRIPTION

ViSP provides simple ways to integrate and validate new algorithms with already existing tools. It follows a module-based software engineering design where data types, algorithms, sensors, viewers and user interaction are made available. Written in C++, ViSP is based on open-source cross-platform libraries (such as OpenCV) and builds with CMake. Several platforms are supported, including OSX, Windows and Linux. ViSP online documentation allows to ease learning. More than 250 fully documented classes organized in 16 different modules, with more than 200 examples and 35 tutorials are proposed to the user. ViSP is released under a dual licensing model. It is open-source with a GNU GPLv2 license. A professional edition license that replaces GNU GPLv2 is also available.

- Participants: François Chaumette, Eric Marchand, Fabien Spindler, Aurélien Yol and Souriya Trinh
- Partner: Inria, Université de Rennes 1
- Contact: Fabien Spindler
- URL: http://visp.inria.fr

In December 2015, ViSP 3.0.0 new modular architecture was released. The corresponding source code tarball was downloaded 2138 times, much more than the previous 2.10.0 release that was downloaded 1412 times. This confirms that ViSP popularity is increasing and motivates the efforts we are doing since more than 10 years to improve the software. ViSP 3.0.0 last release was packaged for Debian, Ubuntu 16.04 LTS, Arch Linux, OSX and ROS. ViSP 3.0.1 next release is in preparation and should be released at the beginning of 2017. This release will be also packaged for iOS devices. In this new version we introduced new wrapper for USB-3 or GigE PointGrey cameras, Haption haptic device, ATI force/torque sensors, Intel RealSense RGB-D devices. We also make an effort to optimize some critical code sections using SSE and make possible cross-compilation for Raspberry PI and iOS targets, and also Nao, Romeo and Pepper robots from SoftBank Robotics. We also introduce a new version of the 3D model-based tracker dedicated to stereo tracking, fixed some issues, improved the documentation by providing new tutorials and by updating the existing ones.

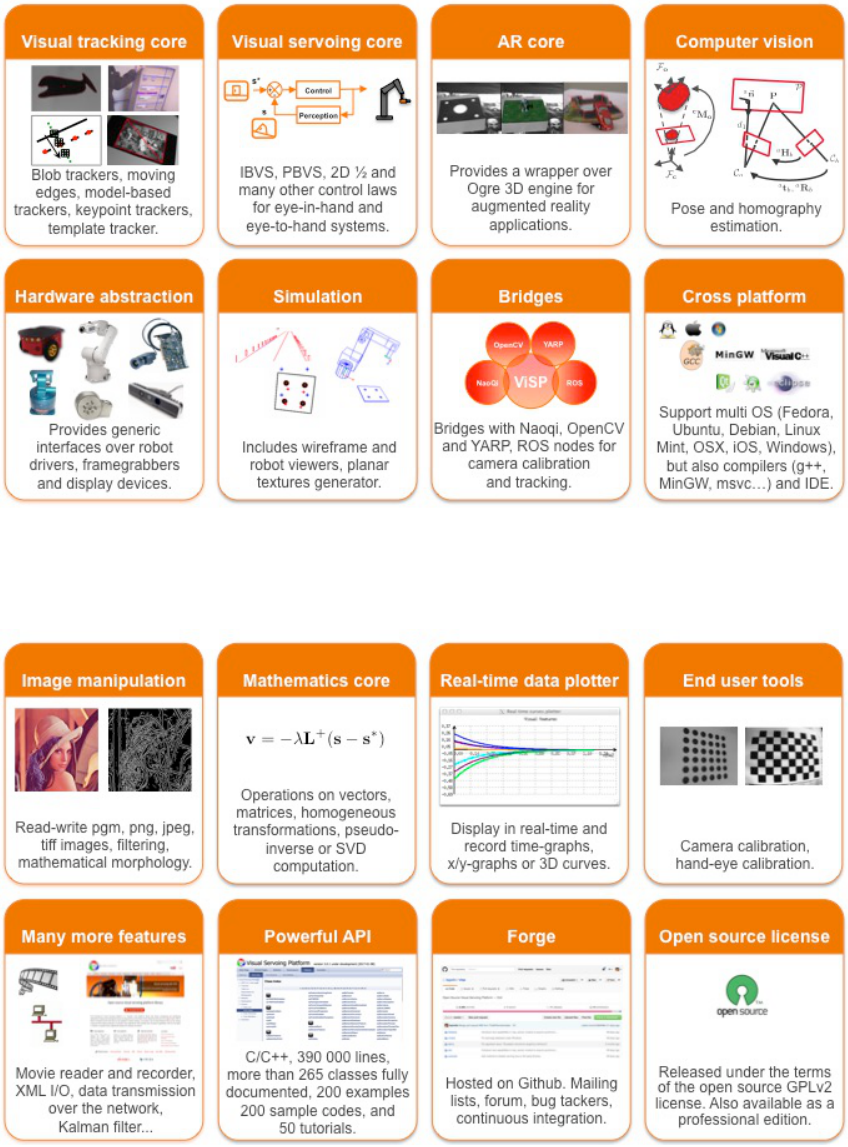


Figure 1. This figure highlights ViSP main capabilities for visual tracking, visual servoing, and augmented reality that may benefit from computer vision algorithms. ViSP allows controlling specific platforms through hardware abstraction or in simulation. ViSP provides also bridges over other frameworks such as OpenCV and ROS. All these capabilities are cross-platform. Moreover, for easing the prototyping of applications, ViSP provides tools for image manipulation, mathematics, data plotting, camera calibration, and many other features. ViSP powerful API is fully documented and available on Github as an open source software under GPLv2 license.

Concerning ROS community, all the existing packages in "`vision_visp`" ROS stack (see http://wiki.ros.org/vision_visp) were updated and ported to kinetic build system. To ease ViSP usage in the ROS framework, the releases of the year were packaged for ROS.

ViSP is used in research labs in France, USA, Canada, Japan, Korea, India, China, Italy, Spain, Portugal, etc. For instance, it is used as a support in graduate courses at IFMA Clermont-Ferrand, University of Picardie in Amiens, Télécom Physique in Strasbourg and ESIR in Rennes. Last August, during the Intel Developer Forum opening keynote, Intel CEO Brian Krzanich introduced the Intel Joule compute module. Using an Intel Joule with glasses from French company PivotHead, Intel demonstrated an augmented reality application that was using ViSP (https://www.youtube.com/watch?v=QRBofzL4MDY).

6.7. WarpDriver

KEYWORDS: Crowd Simulation - Pedestrian Simulation - Collision Avoidance - Reactive Navigation

FUNCTIONAL DESCRIPTION

WarpDriver is a microscopic crowd simulation software, which simulates the collision-free locomotion of many individual agents among the obstacles of a given environment. The originality of the algorithm relies on motion prediction mechanism which allows each agent to predict the probability of colliding other agents with respect to their current motion, their past motion, and the presence of obstacles forcing agents to follow some paths in the environment. Agents then move to their goal whilst they minimize their probability of colliding obstacles.

- Participants: David Wolinski and Julien Pettré
- Contact: Julien Pettré
- URL: http://team.inria.fr/lagadic

6.8. bib2html

FUNCTIONAL DESCRIPTION

The purpose of this software is to automatically produce html pages from BibTEX files, and to provide access to the BibTEX entries by several criteria: year of publication, category of publication, keywords, author name. Moreover cross-linking is generating between pages to provide an easy navigation through the pages without going back to the index.

- Contact: Eric Marchand
- URL: http://www.irisa.fr/lagadic/soft/bib2html/bib2html.html

6.9. Platforms

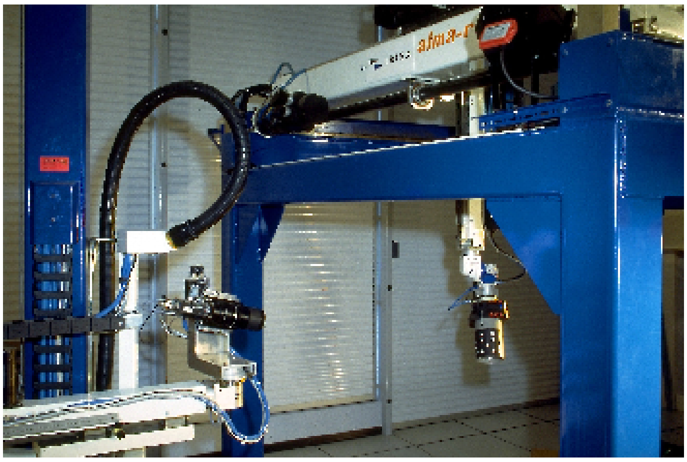
6.9.1. Robot vision platform

Participant: Fabien Spindler [contact].

We exploit two industrial robotic systems built by Afma Robots in the nineties to validate our researches in visual servoing and active vision. The first one is a Gantry robot with six degrees of freedom, the other one is a cylindrical robot with four degrees of freedom (see Fig. 2.a). These robots are equipped with cameras. The Gantry robot also allows embedding grippers on its end-effector.

This year we completed the platform with a haptic Virtuose 6D device from Haption company (see Fig. 2.b). This device is used for visual-based shared control (see Section 9.3.1.3).

Note that 3 papers published by Lagadic in 2016 enclose results validated on this platform [21][48][46].



(a) (b)

Figure 2. a) Lagadic robotics platform for vision-based manipulation, b) Virtuose 6D haptic device

6.9.2. Mobile robots

Participants: Fabien Spindler [contact], Marie Babel, Patrick Rives.

6.9.2.1. Indoor mobile robots

For fast prototyping of algorithms in perception, control and autonomous navigation, the team uses Hannibal in Sophia Antipolis, a cart-like platform built by Neobotix (see Fig. 3.a), and, in Rennes, a Pioneer 3DX from Adept (see Fig. 3.b). These platforms are equipped with various sensors needed for Slam purposes, autonomous navigation and sensor-based control.

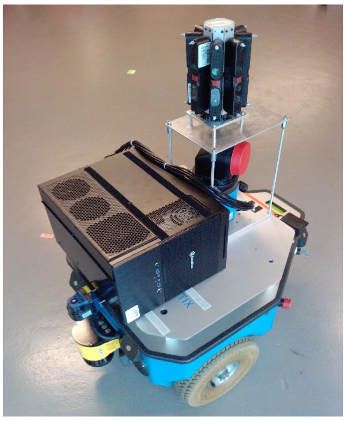
Moreover, to validate the researches in personally assisted living topic (see Section 7.4.5), we have three electric wheelchairs in Rennes, one from Permobil, one from Sunrise and the last from YouQ (see Fig. 3.c). The control of the wheelchair is performed using a plug and play system between the joystick and the low level control of the wheelchair. Such a system lets us acquire the user intention through the joystick position and control the wheelchair by applying corrections to its motion. The wheelchairs have been fitted with cameras and ultrasound sensors to perform the required servoing for assisting handicapped people.

Note that 11 papers exploiting the indoors mobile robots were published this year [67][26][61][37][62][30][55][38][71][64][66].

6.9.2.2. Outdoor vehicles

A camera rig has been developed in Sophia Antipolis. It can be fixed to a standard car (see Fig. 4), which is driven at a variable speed depending on the road/traffic conditions, with an average of 30 km/h and a maximum speed of 80 km/h. The sequences are recorded at a frame rate of 20 Hz, where the six global shutter cameras of the stereo system are synchronized, producing spherical images with a resolution of 2048x665 (see Fig. 4). Such sequences are fused offline to obtain maps that can be used later for localization or for scene rendering (in a similar fashion to Google Street View) as shown in the video http://www-sop.inria.fr/members/Renato-Jose.Martins/iros15.html.

Paper [68] contains experimental results obtained with this camera rig.



(a)

(b)

(c)

Figure 3. a) Hannibal platform, b) Pioneer P3-DX robot, c) wheelchairs from Permobil, Sunrise and YouQ.

Figure 4. Globeye stereo sensor and acquisition system.

6.9.3. Medical robots

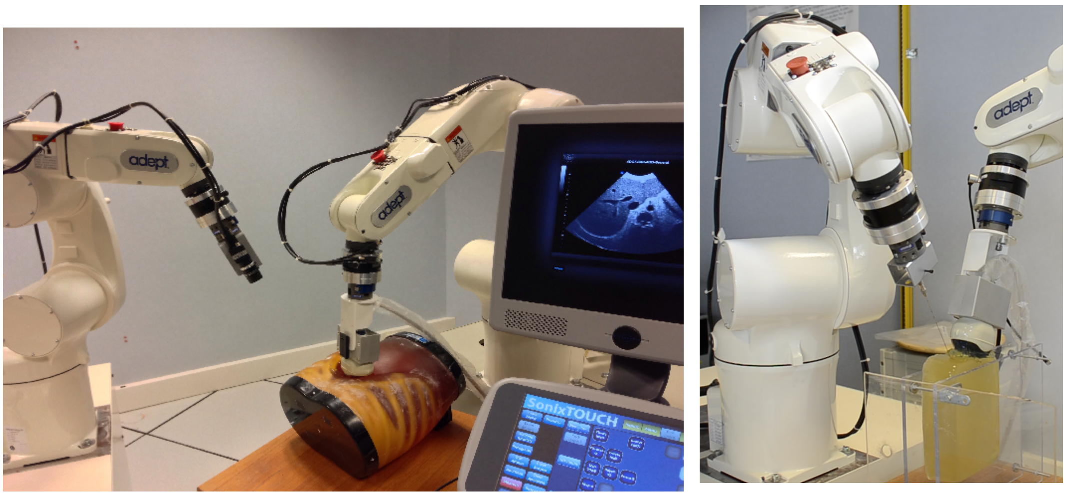
Participants: Fabien Spindler [contact], Alexandre Krupa.

This testbed is of primary interest for researches and experiments concerning ultrasound visual servoing applied to probe positioning, soft tissue tracking or robotic needle insertion tasks (see Section 7.3).

This platform is composed by two Adept Viper six degrees of freedom arms (see Fig. 5.a). Ultrasound probes connected either to a SonoSite 180 Plus or an Ultrasonix SonixTouch imaging system can be mounted on a force torque sensor attached to each robot end-effector.

This year we replaced the F/T sensor attached to one of the Viper robot in order to use a DAQ acquisition board able to provide measures at a higher frame rate (up to 1 kHz). This feature is especially useful for flexible needle steering by ultrasound visual servoing (see Fig. 5.b).

Notice that 10 papers published this year include experimental results obtained with this platform [40][31][34][58][57][70][52][59] [50][51].



(a) (b)

Figure 5. a) Lagadic medical robotics platforms. On the right Viper S850 robot arm equipped with a SonixTouch 3D ultrasound probe. On the left Viper S650 equipped with a tool changer that allows to attach a classical camera or biopsy needles. b) Robotic setup for autonomous needle insertion by visual servoing.

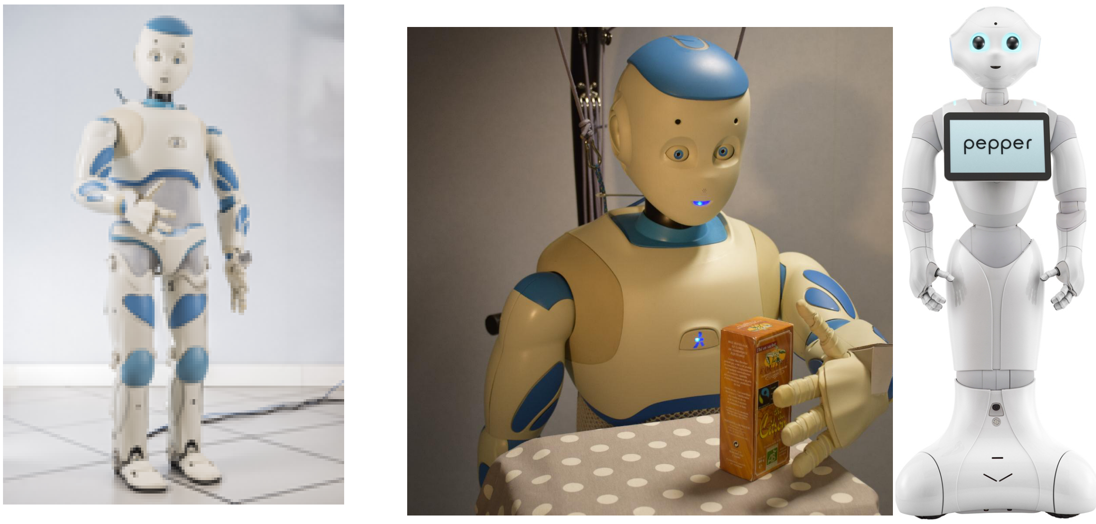
6.9.4. Humanoid robots

Participants: Giovanni Claudio, Fabien Spindler [contact].

Romeo is a humanoid robot from SoftBank Robotics which is intended to be a genuine personal assistant and companion. For the moment only the upper part of the body (trunk, arms, neck, head, eyes) is working. This research platform is used to validate our researches in visual servoing and visual tracking for object manipulation (see Fig. 6.a).

In July, this platform was extended with Pepper, another human-shaped robot designed by SoftBank Robotics to be a genuine day-to-day companion (see Fig. 6.b). It has 17 degrees of freedom mounted on a wheeled holonomic base and a set of sensors (cameras, laser, ultrasound, inertial) that makes this platform interesting for researches in vision-based manipulation and navigation. Our first developments were devoted to visual servoing for controlling the gaze of the robot exploiting the redundancy of the head and mobile base and adding the capability to follow a person.

Note that 4 papers published this year include experimental results obtained with these platforms [53][81][65][20].



(a) (b)

Figure 6. a) Romeo experimental platform, b) Pepper human-shaped robot

6.9.5. Unmanned Aerial Vehicles (UAVs)

Participants: Thomas Bellavoir, Paolo Robuffo Giordano [contact].

From 2014, Lagadic also started some activities involving perception and control for single and multiple quadrotor UAVs, especially thanks to a grant from "Rennes Métropole" (see Section 9.1.5) and the ANR project "SenseFly" (see Section 9.2.5). To this end, we purchased four quadrotors from Mikrokopter Gmbh, Germany (see Fig. 7.a), and one quadrotor from 3DRobotics, USA (see Fig. 7.b). The Mikrokopter quadrotors have been heavily customized by: (i) reprogramming from scratch the low-level attitude controller onboard the microcontroller of the quadrotors, (ii) equipping each quadrotor with an Odroid XU4 board (see Fig. 7.d) running Linux Ubuntu and the TeleKyb software (the middleware used for managing the experiment flows and the communication among the UAVs and the base station), and (iii) purchasing the Flea Color USB3 cameras together with the gimbal needed to mount them on the UAVs (see Fig. 7.c). The quadrotor group is used as robotic platforms for testing a number of single and multiple flight control schemes with a special attention on the use of onboard vision as main sensory modality.

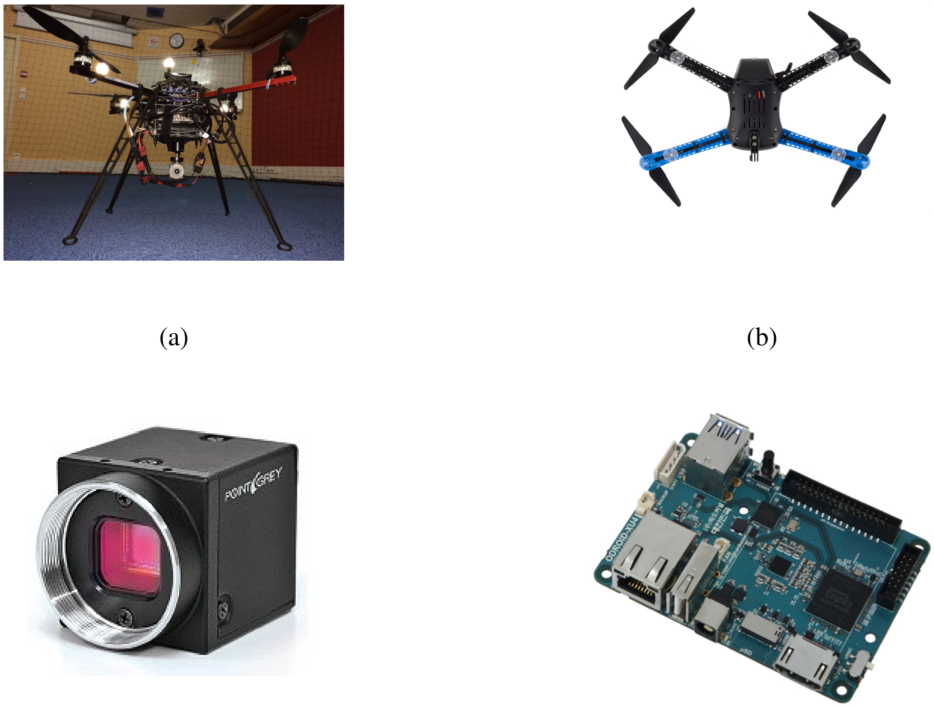
Two papers published this year enclose experimental results obtained with this platform [72][64].

7. New Results

7.1. Visual Perception

7.1.1. Micro/nano Manipulation

Participants: Le Cui, Eric Marchand.



(a) (b) (c) (d)

Figure 7. a) Quadrotor XL1 from Mikrokopter, b) Quadrotor Iris from 3DRobotics, c) Flea Color USB3 camera, d) Odroid XU4 board

Le Cui's Ph.D. [15] ended with a contribution related to visual tracking and estimation of the 3D pose of a micro/nano-object. It is indeed a key issue in the development of automated manipulation tasks using visual feedback. The 3D pose of the micro object can be estimated based on a template matching algorithm. Nevertheless, a key challenge for visual tracking in a scanning electron microscope (SEM) was the difficulty to observe the motion along the depth direction. We then proposed a template-based hybrid visual tracking scheme that uses luminance information to estimate the object displacement on x-y plane and uses defocus information to estimate object depth [54].

7.1.2. 3D Localization for Space Debris Removal

Participants: Aurélien Yol, Eric Marchand, François Chaumette.

This study is realized in the scope of the FP7 Removedebris project (see Section 9.3.1.1) [27]. We compared two vision-based navigation methods for tracking space debris in a low Earth orbit environment. The proposed approaches rely on a frame to frame model-based tracking in order to obtain the complete 3D pose of the camera with respect to the target [2]. The proposed algorithms robustly combine points of interest and edge features, as well as color-based features if needed. Experimental results have been obtained demonstrating the robustness of the approaches on synthetic image sequences simulating a CubeSat satellite orbiting the Earth [75].

7.1.3. 3D Localization for Airplane Landing

Participants: Noël Mériaux, François Chaumette, Patrick Rives, Eric Marchand.

This study is realized in the scope of the ANR VisioLand project (see Section 9.2.2). In a first step, we have considered and adapted our model-based tracker [2] to localize the aircraft with respect to the airport surroundings. Satisfactory results have been obtained from real image sequences provided by Airbus. In a second step, we are now considering to perform this localization from a set of keyframe images corresponding to the landing trajectory.

7.1.4. Scene Registration based on Planar Patches

Participants: Renato José Martins, Eduardo Fernandez Moral, Patrick Rives.

Image registration has been a major problem in computer vision over the past decades. It implies searching an image in a database of previously acquired images to find one (or several) that fulfill some degree of similarity, e.g. an image of the same scene from a similar viewpoint. This problem is interesting in mobile robotics for topological mapping, re-localization, loop closure and object identification. Scene registration can be seen as a generalization of the above problem where the representation to match is not necessarily defined by a single image (i.e. the information may come from different images and/or sensors), attempting to exploit all information available to pursue higher performance and flexibility. This problem is ubiquitous in robot localization and navigation. We propose a probabilistic framework to improve the accuracy and efficiency of a previous solution for structure registration based on planar representation. Our solution consists of matching graphs where the nodes represent planar patches and the edges describe geometric relationships. The maximum likelihood estimation of the registration is estimated by computing the graph similarity from a series of geometric properties (areas, angles, proximity, etc.) to maximize the global consistency of the graph. Our technique has been validated on different RGB-D sequences, both perspective and spherical [26].

7.1.5. Direct RGB-D Registration

Participants: Renato José Martins, Eduardo Fernandez Moral, Patrick Rives.

Dense direct RGB-D registration methods are widely used in tasks ranging from localisation and tracking to 3D scene reconstruction [7]. This work addresses a peculiar aspect which drastically limits the applicability of direct registration, namely the weakness of the convergence domain. In general, registration is performed only between close frames (small displacements), since dense registration tasks are particularly sensible to the local convexity of the cost error function. The main contribution of this work is an adaptive RGB-D error cost function that has a larger convergence domain and a faster convergence in both simulated and real data

[67], [68]. This formulation employs the relative condition number metric to update the weighting of the RGB and depth costs. This approach is performed within a multi-resolution framework, where an efficient pixel selection for both RGB and ICP costs reduces the computational cost whilst preserving the precision. The formulation results in a larger region of attraction and faster convergence than classical RGB, ICP and RGB-D costs. Experiments was conducted using real sequences of indoor and outdoor images using perspective and spherical RGB-D sensors. Significant improvements was denoted in terms of the convergence stability and the speed of convergence in comparison with state-of-the-art methods.

7.1.6. Online localization and mapping for UAVs

Participants: Muhammad Usman, Paolo Robuffo Giordano, Eric Marchand.

Localization and mapping in unknown environments is still an open problem, in particular for what concerns UAVs because of the typical limited memory and processing power available onboard. In order to provide our quadrotor UAVs with high autonomy, we started studying how to exploit onboard cameras for an accurate (but fast) localization and mapping in unknown indoor environments. We chose to base both processes on the newly available Semi-Direct Visual Odometry (SVO) library (http://rpg.ifi.uzh.ch/software) which has gained considerable attention over the last years in the robotics community. The idea is to exploit dense images (i.e., with little image pre-processing) for obtaining an incremental update of the camera pose which, when integrated over time, can provide the camera localization (pose) w.r.t. the initial frame. In order to reduce drifts during motion, a concurrent mapping thread is also used for comparing the current view with a set of keyframes (taken at regular steps during motion) which constitute a “map” of the environment. We have started porting the SVO library to our UAVs and the preliminary results showed good performance of the localization accuracy against the Vicon ground truth. We are now planning to close the loop and base the UAV flight on the reconstructed pose from the SVO algorithm.

7.1.7. Reflectance and Illumination Estimation for Realistic Augmented Reality

Participants: Salma Jiddi, Eric Marchand.

The acquisition of surface material properties and lighting conditions is a fundamental step for photo-realistic Augmented Reality. Human visual cues remain sensitive to the global coherence within a computer-generated image. Absence or bad rendered virtual shadows, unconsidered specular reflections and/or occlusions, confused color perception such as an exuberantly bright virtual object are all elements which may not help an AR user interact and commit to a target application. In this work, we studied a new method for the estimation of the diffuse and specular reflectance properties of an indoor real static scene. Using an RGB-D sensor, we further estimate the 3D position of light sources responsible for specular phenomena and propose a novel photometry-based classification for all the 3D points. The resulting algorithm allows convincing AR results such as realistic virtual shadows as well as proper illumination and specularity occlusions [60].

7.1.8. Optimal Active Sensing Control

Participants: Paolo Salaris, Riccardo Spica, Paolo Robuffo Giordano.

This study concerns the problem of active sensing control. The objective is to improve the estimation accuracy of an observer by determining the inputs of the system that maximize the amount of information gathered by the outputs. In [9] this problem has been solved within the Structure from Motion (SfM) framework for 3D structure estimation problems, i.e. a point, a sphere and a cylinder, in the particular case where the observability property is instantaneously guaranteed. The optimal estimation strategy is hence given in terms of the instantaneous velocity direction of the camera velocities.

Recently, we have extended the optimal active sensing control to the case where the observability property is not instantaneously guaranteed. To simplify the analysis, we considered nonlinear differentially flat systems. Moreover, to quantify the richness of the acquired information the Observability Gramian (OG) has been used. We have hence defined a trajectory for the flat outputs of the system by using B-Spline curves and then, we have exploited an online gradient descent strategy to move the control points of such B-Spline in order to actively maximise the smallest eigenvalue of the OG over the whole fixed planning time horizon. While the system travels along its planned (optimized) trajectory, an Extended Kalman Filter (EKF) is used to estimate the system state. In order to keep memory of the past acquired sensory data for online re-planning, the OG is also computed on the past estimated state trajectories. This is then used for an online replanning of the optimal trajectory during the robot motion which is continuously refined by exploiting the estimated system state by the EKF. In order to show the effectiveness of our method we have considered a simple but significant case of a planar robot with a single range measurement. The simulation results show that, along the optimal path, the EKF converges faster and provides a more accurate estimate than along any other possible (non-optimal) paths. These results have been submitted to ICRA'2017.

7.2. Sensor-based Robot Control

7.2.1. Determining Singularity Configurations in IBVS

Participant: François Chaumette.

This theoretical study has been achieved through an informal collaboration with Sébastien Briot and Philippe Martinet from IRCCyN in Nantes, France. It concerns the determination of the singularity configurations of image-based visual servoing using tools from the mechanical engineering community and the concept of "hidden" robot. In a first step, we have revisited the welknown case of using three image points as visual feature, and then solved the general case of n image points [22]. The case of three image straight lines has also been solved for the first time [23].

7.2.2. Interval-based IBVS convergence domain computation

Participant: Vincent Drevelle.

This work aims to compute the set of camera poses from which IBVS will converge to the desired pose (that corresponds to the reference image). Starting from a (small) initial attraction domain of the desired pose (obtained using Lyapunov theory), we employ subpavings and guaranteed integration to iteratively increase the proven convergence domain, using a viability-based approach. Image-domain and pose-domain constraints are also enforced, like feature points visibility or workspace boundaries. First results have been obtained for a 3DOF line-scan camera IBVS case [56].

7.2.3. Visual Servoing of Humanoid Robots

Participants: Giovanni Claudio, Don Joven Agravante, Fabien Spindler, François Chaumette.

This study is realized in the scope of the BPI Romeo 2 and H2020 Comanoid projects (see Sections 9.2.7 and 9.3.1.2).

In a first step, we have considered classical kinematic visual servoing schemes for gaze control and manipulation tasks, such as can or box grasping. Two-hand manipulation has also been achieved using a master/slave approach [53], [81]. In a second step, we have designed the modeling of the visual features at the acceleration level to embed visual tasks and visual constraints in an existing QP controller [20][80]. Experimental results have been obtained on Romeo (see Section 6.9.4).

7.2.4. Model Predictive Visual Servoing

Participants: Nicolas Cazy, Paolo Robuffo Giordano, François Chaumette.

This study is realized in collaboration with Pierre-Brice Wieber, from Bipop group at Inria Rhône Alpes.

Model Predictive Control (MPC) is a powerful control framework able to take explicitly into account the presence of constraints in the controlled system (e.g., actuator saturations, sensor limitations, and so on). In this research activity, we studied the possibility of using MPC for tackling one of the most classical constraints of visual servoing applications, that is, the possibility to lose tracking of features because of occlusions, limited camera field of view, or imperfect image processing/tracking. The MPC framework depends upon the possibility to predict the future evolution of the controlled system over some time horizon, for correcting the current state of the modeled system whenever new information (e.g., new measurements) become available. We have also explored the possibility of applying these ideas in a multi-robot collaboration scenario where a UAV with a downfacing camera (with limited field of view) needs to provide localization services to a team of ground robots [13].

7.2.5. Model Predictive Control for Visual Servoing of a UAV

Participants: Bryan Penin, Riccardo Spica, François Chaumette, Paolo Robuffo Giordano.

Visual servoing is a welknown class of techniques meant to control the pose of a robot from visual input by considering an error function directly defined in the image (sensor) space. These techniques are particularly appealing since they do not require, in general, a full state reconstruction, thus granting more robustness and lower computational loads. However, because of the quadrotor underaction and inherent sensor limitations (mainly limited camera field of view), extending the classical visual servoing framework to the quadrotor flight control is not straightforward. For instance, for realizing a horizontal displacement the quadrotor needs to tilt in the desired direction. This tilting, however, will cause any downlooking camera to point in the opposite direction with, e.g., possible loss of feature tracking because of the limited camera field of view.

In order to cope with these difficulties and achieve a high-performance visual servoing of quadrotor UAVs, we are exploring the possibility of using techniques borrowed from Model-Predictive Control (MPC) for explicitly dealing with this kind of constraints during flight. Indeed, MPC is a class of (numerical) optimal control techniques able to explicitly take into account state and input constraints, as well as complex (and underactuated) nonlinear dynamics of the controlled system. In particular, the ability to predict, over some future time window, the behavior of the visual features on the image plane will allow the quadrotor to fly “blindly” for some limited phases, for then regaining tracking of any lost feature. This possibility will be crucial for allowing quick maneuvering guided by a direct visual feedback. We have started addressing the case of a simulated planar UAV as a representative case study, and we are now working towards an experimental validation with a real quadrotor UAV equipped with an onboard camera.

7.2.6. Visual-based shared control

Participants: Firas Abi Farraj, Nicolò Pedemonte, Paolo Robuffo Giordano.

This work concerns our activities in the context of the RoMaNS H2020 project (see Section 9.3.1.3. Our main goal is to allow a human operator to be interfaced in an intuitive way with a two-arm system, one arm carrying a gripper (for grasping an object), and the other one carrying a camera for looking at the scene (gripper + object) and providing the needed visual feedback. The operator should be allowed to control the two-arm system in an easy way for letting the gripper approaching the target object, and she/he should also receive force cues informative of how feasible her/his commands are w.r.t. the constraints of the system (e.g., joint limits, singularities, limited camera fov, and so on).

We have started working on this topic by proposing a shared control architecture in which the operator could provide instantaneous velocity commands along four suitable task-space directions not interfering with the main task of keeping the gripper aligned towards the target object (this main task was automatically regulated). The operator was also receiving force cues informative of how much her/his commands were conflicting with the system constraints, in our case joint limits of both manipulators. Finally, the camera was always moving so as to keep both the gripper and the target object at two fixed locations on the image plane [46].

We have then extended this framework in two directions: first, by allowing the possibility of controlling a whole future trajectory for both arms (gripper+camera) while coping with the system constraints. The operator was then receiving an ‘integral’ force feedback along the whole planned trajectory: in this way, the operator’s actions and the corresponding force cues were function of a planned trajectory (thus, carrying information over a future time window) that could be manipulated at runtime. Second, we studied how to integrate learning from demonstration into our framework by first using learning techniques for extracting statistical regularities of ‘expert users’ executing successful trajectories for the gripper towards the target object. Then, these learned trajectories were used for generating force cues able to guide novice users during their teleoperation task by the ‘hands’ of the expert users who demonstrated the trajectories in the first place. Both works have been submitted to ICRA’2017.

7.2.7. Direct Visual Servoing

Participants: Quentin Bateux, Eric Marchand.

In the direct visual servoing methods such as photometric framework, the images as a whole are used to define the control law. This can be opposed to the classical visual servoing approaches that relies on geometric features and where image processing algorithms that extract and track visual features are necessary. In [21], we proposed a generic framework to consider histograms as visual features. A histogram is an estimate of the probability distribution of a variable (for example the probability of occurrence in an intensity, color, or gradient orientation in an image). We demonstrated that the framework we proposed applies, but is not limited to, a wide set of histograms and allows the definition of efficient control laws.

Nevertheless, the main drawback for the direct visual servoing class of methods comparing to the classical geometric visual servoing methods is their comparatively limited convergence range. We then proposed in [48] a new direct visual servoing control law that relies on a particle filter to perform non-local and non-linear optimization in order to increase the convergence domain. To each particle considered we associate a virtual camera that predicts the image it should capture by using image transfer techniques. This new control law has been validated on a 6 DOF positioning task performed on our Gantry robot (see Section 6.9.1).

7.2.8. *Audio-based Control*

Participants: Aly Magassouba, François Chaumette.

This study is concerned with the application of sensor-based control approach to audio sensors. It is made in collaboration with Nancy Bertin from Panama group at Irisa and Inria Rennes-Bretagne Atlantique. Auditory features such as Interaural Time Difference (ITD), Interaural Level Difference (ILD), and sound energy have been modeled and integrated in various control schemes to control the motion of a mobile robot with two microphones onboard [66], [64]. Experiments with Romeo and Pepper (see Section 6.9.4) have also been achieved [65]. They show the robustness of closed loop sensor-based control with respect to coarse modeling and that explicit sound source localization is not a mandatory step for aural servoing.

7.3. Medical Robotics

7.3.1. *Non-rigid Target Tracking in Ultrasound Images*

Participants: Lucas Royer, Alexandre Krupa.

We pursued our work concerning the development of a real-time approach that allows tracking deformable soft tissue structures in 3D ultrasound sequences. In previous work we proposed a method which consists in estimating the target deformation by combining robust dense motion estimation and mechanical model simulation. This year we improved the robustness of our method to several image artefacts as the presence of large shadows, local illumination changes and image occlusions that occur due to the modification of the imaging gain and re-orientation of the ultrasound beam induced by probe motion. To achieve this, we proposed a new dissimilarity criterion between the current and reference images based on the Sum of Conditional Variance (SCV). Our new criterion, that we named Sum of Confident Conditional Variance (SCCV), consists in discriminating unconfident voxels thanks to the use of a pixel-wise quality measurement of the ultrasound images. This improved approach was experimentally validated on organic soft tissues and the obtained results were published in [40].

7.3.2. *Optimization of Ultrasound Image Quality by Visual Servoing*

Participants: Pierre Chatelain, Alexandre Krupa.

This study is realized in collaboration with Prof. Nassir Navab from the Technical University of Munich (TUM).

In previous work, we have developed ultrasound-based visual servoing methods to fulfill various tasks, such as compensating for physiological motion, maintaining the visibility of an anatomic target during ultrasound probe teleoperation, or tracking a surgical instrument. However, due to the specific nature of ultrasound images, guaranteeing a good image quality during the procedure remains a challenge. Therefore we pursued our study on the use of ultrasound confidence maps as a new modality for automatically positioning an ultrasound probe in order to improve the image quality. In addition to our visual servoing approach that optimizes the global quality of the image, this year we proposed a control fusion to optimize the acoustic window for a specific anatomical target which is tracked in the ultrasound images [50]. Recently, we extended our confidence-driven control to the out-of-plane motion of a 3D ultrasound probe and experimentally validated it on a human volunteer at TUM [14].

7.3.3. Visual Servoing using Shearlet Transform

Participants: Lesley-Ann Duflot, Alexandre Krupa.

In collaboration with the Femto-ST lab in Besançon, we proposed in a first-hand a solution to reduce the acquisition time of an Optical Coherence Tomography (OCT) 3D imaging scanner. This latter consists in sweeping a laser beam on a tissue sample of interest. To increase the frame rate of this imaging device we proposed to apply an optimal trajectory to the laser that covers entirely the image but without performing all the OCT measurements. The reconstruction of the missing data is then achieved by applying an updated Fast Iterative Soft-Thresholding Algorithm (FISTA) on a sparse representation of the image that is based on the shearlet transform [57]. In a second hand, we studied the feasibility of using the subsampled shearlet coefficients of an ultrasound image as the visual features of an image-based visual servoing. In a preliminary study we estimated numerically the interaction matrix that links the variation of the shearlet coarsest coefficients to the 6 degrees of freedom motion of the ultrasound probe and uses it in the visual servoing framework. The results obtained in cases of automatic probe positioning and phantom motion compensation demonstrated the efficiency of the shearlet-based features in terms of accuracy, repeatability, robustness and convergence behavior [59]. Then we proposed to consider a more efficient and adequate shearlet implementation that consists in a non-subsampled representation of the image. In this case the shearlet coefficients represent different images, focused on different singularities of the initial image, and we consider directly their pixel intensity values in the visual feature vector similarly to the photometry-based visual servoing approach. The modeling of the interaction matrix was analytically derived and experimental results demonstrated the reliability of the new method and its robustness to speckle noise [58].

7.3.4. 3D Steering of Flexible Needle by Ultrasound Visual Servoing

Participants: Jason Chevrie, Marie Babel, Alexandre Krupa.

The objective of this work is to provide robotic assistance during needle insertion procedures such as biopsy or ablation of localized tumor. In the past we only considered the control of the insertion and needle rotation along and around its main axis by the use of a duty-cycling control strategy. This latter consists in adapting online from visual feedback the orientation of a beveled-tip flexible needle during its insertion for controlling the needle curvature in 3D space that is induced by asymmetrical forces exerted on the bevel. However, such strategy limits the workspace of the needle tip. Therefore we proposed a new control method for flexible needle steering that combines direct base manipulation and needle tip based control. The direct base manipulation control is generated thanks to the use of a 3D model of a flexible beveled tip needle that gives the adequate motion of the needle base to obtain a given motion of the needle tip. This 3D model is based on virtual springs that characterize the needle mechanical interaction with soft tissue and is adapted online from visual tracking of the needle shape. From this model, a measure of the controllability of the needle tip degrees of freedom was proposed in order to mix the control between the direct base manipulation and the duty cycling technique [51]. Preliminary results of an automatic needle tip positioning in a translucent gelatine phantom, observed by 2 orthogonal cameras, demonstrated the feasibility of the combination between direct base manipulation and needle tip control for reaching a desired target. This hybrid control allows better targeting capabilities in terms of larger needle workspace and reduced needle bending. In order to predict the trajectory of a needle during insertion under lateral motion of the tissue, we also improved our 3D model of the flexible needle to take into account the effect of the motion of the tissues on the needle shape. This was achieved thanks to the design of an algorithm based on an unscented Kalman filter that estimates the tissue motion. Results obtained from several needle insertions in a moving soft tissue phantom showed that our model gives good performance in terms of needle trajectory prediction. This model was also considered in a closed-loop control approach to allow automatic reaching of a target in case of tissue lateral displacement [52]. Future work will address the consideration of 3D ultrasound as visual feedback.

7.3.5. Enhancement of Ultrasound Elastography by Visual Servoing and Force Control

Participants: Pedro Alfonso Patlan Rosales, Alexandre Krupa.

Elastography imaging is performed by applying continuous stress variation on soft tissues in order to estimate a strain map of the observed tissues. It is obtained by estimating, from the RF (radio-frequency) signal along each scan line of the probe transducer, the echo time delays between pre- and post-compressed tissue. Usually, this continuous stress variation is performed manually by the user who manipulates the US probe and it results therefore in an user-dependent quality of the elastography image. To improve the ultrasound elastography imaging and provide quantitative measurement, we developed an assistant robotic palpation system that automatically moves a 2D ultrasound probe for optimizing ultrasound elastography [70]. The main originality of this work is the use of the elastography modality directly as input of the robot controller. Force measures are also considered in the probe control in order to automatically induce soft tissue deformation needed for real-time elastography imaging process.

7.4. Navigation of Mobile Robots

7.4.1. Visual Navigation from an Image Memory

Participants: Suman Raj Bista, Paolo Robuffo Giordano, François Chaumette.

This study is concerned with visual autonomous navigation in indoor environments. As in our previous works concerning navigation outdoors [4], the approach is based on a topological localization of the current image with respect to a set of keyframe images, but the visual features used for this localization as well as for the visual servoing are not composed of points of interest, but either on mutual information [71] following the idea proposed in [3], or straight lines that are more common indoors [38], or finally on a combination of points of interest and straight lines [11]. Satisfactory experimental results have been obtained using the Pioneer mobile robot (see Section 6.9.2).

7.4.2. Robot-Human Interactions during Locomotion

Participant: Julien Pettré.

In collaboration with the Gepetto team of Laas in Toulouse and the Mimetic group in Rennes, we have studied how humans avoid collision with a robot. Understanding how humans achieve such avoidance is crucial to better anticipate humans' reactions to the presence of a robot and to control the robot to adapt its trajectory accordingly. It is generally assumed that humans avoid a robot just like they avoid another human. In this work, we bring the empirical demonstration that humans actually set a specific strategy to avoid robots, and that, more precisely, they show a preference to give way to a robot which is on a collision course with them [41]. This results brings useful insight about human-robot interactions during locomotion, and provides useful guidelines to design reactive navigation techniques for mobile robots aimed at moving among humans.

7.4.3. Scene Mapping based on Intelligent Human/Robot Interactions

Participant: Patrick Rives.

For mobile robots to operate in compliance with human presence, interpreting the impact of human activities and responding constructively is a challenging goal. Towards this objective, mapping an environment allows robots to be deployed in diverse workspaces, marking this skill as a primary element in the integration of robots into human-populated environments. We proposed an effective approach for using human activity cues in order to enhance robot mapping and navigation and in particular in filtering noisy human detections, detecting passages, inferring space occupancy and allowing navigation within unexplored areas. Our contributions [36] are based on the development of intelligent interactions among conceptually different mapping levels, namely, the metric, social and semantic levels. Experiments, using the Hannibal platform (see Section 6.9.2), highlighted a number of strong dependences among these levels and the way in which they can be used to enhance individual performances and in turn the global robot operation.

7.4.4. Autonomous Social Navigation of a Wheelchair

Participants: Vishnu Karakkat Narayanan, Marie Babel.

This work is realized in collaboration with Anne Spalanzani (Chroma team - Inria Grenoble).

A key issue that hinders the adoption of assistive robotic technologies such as robotized wheelchair, in the real world, is that they need to operate in mostly human environments and among human crowds. Indeed intelligent wheelchairs need to be deployed in a human environment thereby making it essential for such robots to incorporate a sense of human-awareness. Simply put, humans are special objects that have to be perceived and acted on in a special manner by robots that interact with us humans. Thus one can define Human-aware Navigation as an intersection between human-robot interaction and robotic motion planning.

In this context we introduced a low-level velocity controller that could be employed by a social robot like a robotic wheelchair for approaching a group of interacting humans, in order to become a part of the interaction. Taking into account an interaction space that is created when at least two humans interact, a meeting point can be calculated where the robot should reach in order to equitably share space among the interacting group. We then proposed a sensor-based control law which uses the position and orientation of the humans with respect to the sensor as inputs, to reach the meeting point while respecting spatial social constraints [61]. Experiments using a mobile robot equipped with a single laser scanner, realized in collaboration with Ren Luo (Taiwan) within the Sampen Inria associated team, also proved the success of the algorithm in a noisy real world scenario [62].

In addition, a semi-autonomous framework for human-aware navigation in an intelligent wheelchair has been designed. A generalized linear control sharing framework was proposed that was able to progressively correct the user teleoperation in order to avoid obstacles and in order to avoid disturbance to humans. Meanwhile, we also proposed a Bayesian approach for user intention estimation. The formulation was partly inferred from the design of the controller for assisted doorway passing, wherein we hypothesized that predicting short term goals is sufficient for eliminating user intention uncertainty [16].

7.4.5. Semi-autonomous Control of a Wheelchair for Navigation Assistance

Participants: Louise Devigne, Vishnu Karakkat Narayanan, Marie Babel.

To address the wheelchair driving assistance issue, we proposed a unified shared control framework able to smoothly correct the trajectory of the electrical wheelchair [16]. The system integrates the manual control with sensor-based constraints by means of a dedicated optimization strategy. The resulting low-complex and low-cost embedded system is easily plugged onto on-the-shelf wheelchairs [55]. The robotic solution is currently under validation process with volunteering patients of Pôle Saint Hélier (France) who present different disabling neuro-pathologies preventing them to drive non-assisted wheelchairs.

Within the frame of ISI4NAVE associated team (see Section 9.4.1.2), this shared-control solution has been then coupled with first experimental biofeedback devices such as haptic devices. Preliminary tests have been conducted within the PAMELA facility at University College of London and within the rehabilitation center of Pôle Saint Hélier in Rennes (see Section 8.1.3). They involved regular wheelchair users as well as medical staff. We have demonstrated the ability of the framework to provide relevant assistance and now need to focus on methods to fine-tune parameters and customize/calibrate to the individual and evolving needs of each user.

7.5. Multi-robot and Crowd Motion Control

7.5.1. Advanced multi-robot control and estimation

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 fulfillment 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. We gave two contributions in this field: in the first one [35], we proposed a decentralized exploration strategy for a team of 3D agents able to guarantee exploration of a finite space in a finite amount of time while coping with the constraints of a connected sensing/communication graph for the robot group against sensing/communication constraints (limited range, occluded line-of-sight), and of obstacle and inter-robot collision avoidance. The strategy exploits a suitable state machine for assigning dynamic roles to the agents in the group for allowing completion of the exploration in finite time. Second, in [28] we studied how the choice of a leader agent in a leader-follower scenario could affect the performance of the group when tracking a desired formation (shape and gross motion). The proposed strategy allows selecting the "best leader" online as a function of the current group state (relative positions and velocities) and of the group topology (assumed connected). By cycling among several connected topologies during motion, we could show that our proposed leader selection algorithm provides the best performance among other possible choices (including the random one) while coping with the constraint of a connected (but possibly time-varying) topology.

These works were realized in collaboration with the robotics group at the Max Planck Institute for Biological Cybernetics, Tübingen, Germany, and the RIS group at Laas in Toulouse.

7.5.2. Rigidity-based methods for formation control

Participants: Fabrizio Schiano, Riccardo Spica, Andrea Peruffo, Paolo Robuffo Giordano.

Most multi-robot applications must rely on *relative sensing* among the robot pairs (rather than absolute/external sensing such as, e.g., GPS). For these systems, the concept of *rigidity* provides the correct framework for defining an appropriate sensing and communication topology architecture. Rigidity is a combinatorial theory for characterizing the "stiffness" or "flexibility" of structures formed by rigid bodies connected by flexible linkages or hinges. In a broader context, rigidity turns out to be an important architectural property of many multi-agent systems when a common inertial reference frame is unavailable. Applications that rely on sensor fusion for localization, exploration, mapping and cooperative tracking of a target, all can benefit from notions in rigidity theory. The concept of rigidity, therefore, provides the theoretical foundation for approaching decentralized solutions to the aforementioned problems using distance measurement sensors, and thus establishing an appropriate framework for relating system level architectural requirements to the sensing and communication capabilities of the system.

In the recent past, we have proposed a decentralized gradient-based rigidity maintenance action for a group of quadrotor UAVs [10]. By starting in a rigid configuration, the group of UAVs was 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. This (rigidity-based) control/estimation framework has now been extended to the case of *bearing rigidity* for directed graphs: here, rather than distances the measurements are the 3D bearing vectors expressed in the local body-frame of each agent. The theory has been extended to the case of 3D agents evolving in $\mathbb{R}^3 \times \mathbb{S}^1$ by proposing a decentralized bearing controller/localization algorithm that only requires one single distance measurement (among an arbitrary pair of agents) for a correct convergence [72]. The proposed algorithm ensures stabilization towards a desired bearing formation, and allows for the possibility of actuating the motion directions in the null-space of the bearing constraints (that is, collective translations in 3D, expansion/retraction, and coordinated rotation about a vertical axis).

The need of a single distance measurement (for fixing the formation scale) has also been relaxed in [73] where an *active* scale estimation scheme has been proposed for allowing the (distributed) estimation of the various inter-agent distances online by processing the measured bearings and the known agent ego-motion (body-frame linear and angular velocities). Finally, we have also proposed an extension of the "distance" rigidity maintenance controller proposed in [10] to the case of bearing measurements (and bearing rigidity), by considering the typical sensing constraints of onboard cameras, that is, limited range, limited field of view, of possible mutual occlusions when two or more agents lie on the same line-of-sight. This work has been experimentally validated with 5 quadrotor UAVs, and has been submitted to ICRA'2017.

These works were realized in collaboration with the RIS group at Laas, Toulouse, and with Technion, Israel.

7.5.3. Cooperative localization using interval analysis

Participants: Ide Flore Kenmogne Fokam, Vincent Drevelle.

In the context of multi-robot fleets, cooperative localization consists in gaining better position estimate through measurements and data exchange with neighboring robots. Positioning integrity (i.e., providing reliable position uncertainty information) is also a key point for mission-critical tasks, like collision avoidance. The goal of this work is to compute position uncertainty volumes for each robot of the fleet, using a decentralized method (i.e., using only local communication with the neighbors). The problem is addressed in a bounded-error framework, with interval analysis and constraint propagation methods. These methods enable to provide guaranteed position error bounds, assuming bounded-error measurements. They are not affected by over-convergence due to data incest, which makes them a well sound framework for decentralized estimation. Ongoing work focuses on position uncertainty domain computation in image-based UAV localization [63], and its extension to cooperative localization in a multi-UAV fleet.

7.5.4. Numerical Models of Local Interactions during Locomotion

Participants: Julien Bruneau, Panayiotis Charalambous, David Wolinski, Julien Pettré.

The numerical models of local interactions are core components of reactive navigation techniques (which allows a robot to avoid dynamic obstacles) and of microscopic crowd simulation algorithms (which allows to simulate a crowd motion as a collection of agent trajectories). We have pursued our efforts to design local models of interactions which capture humans pedestrian behavior, to simulate how they adapt their trajectory so as to perform interactions with their neighbors [12]. This year, our efforts were focused on the simulation of grouping behaviors [39], and mid-term strategies human set to perform energy-efficient sequences of successive avoidance adaptations [24]. These two situations deal with complex situations of interactions, where several interactions of different kinds need to be combined to compute agents trajectories. For example, when moving in groups, agents have to keep close to the other members of their group while they should not collide with them, as well as they should avoid collision with any other agent or obstacle out of this group.

We also revisited the foundation of velocity-based models of local interaction for collision avoidance. Using a velocity-based model, a collision-free motion is computed for one agent by extrapolating the future motion of neighbor agents with respect to their current position and velocity. From this information, each agent can deduce the set of velocities (called admissible velocities) that lead to a collision-free motion in the near future. The extrapolation is generally simply based on a linear extrapolation of the future position along the current velocity vector. This is simplistic as it assumes that the current velocity vector is representative of the future motion, while it is often false when, for instance, the agent is currently performing adaptations due to ongoing collision avoidance, or when the agent is following a curvy path. To improve the accuracy of motion prediction and the resulting simulation, we have introduced a probabilistic representation of future position, that can be computed from a set of context elements such as the layout of the environment or the agents past motion [42]. We demonstrate in this work the high impact on the level of realism of resulting simulations. This work is implemented in the WarpDriver software (see Section 6.7).

Finally, we address applications of our simulators to the Computer Animation. Crowd simulation agents generally have a simplistic geometrical and kinematics models, typically, an oriented 2D circle moving on a flat surface. In Computer Animation, an animation of a crowd of 3D realistic characters can be computed on top of the agents simulation by computing their internal joints trajectories so as to perform walking motion along computed agents trajectories. However, the discrepancies between the 2D model of agents and 3D full body characters may result into residual collisions between character shapes. In this collaboration with the Mimetic team, we demonstrate that simple secondary animations for characters, such as local shoulder motions, can be efficiently triggered to camouflage those artefacts, with a very low computational overhead [29].

7.5.5. Motion Planning for Digital Characters

Participant: Julien Pettré.

Motion planning is an important component for agents and robot navigation and control, providing them the ability to perform geometrical reasoning over their environment to transform a high-level distant goal in their environment into a sequence of local motions and sub-goals to reach. This year, we have been involved into two collaborations dealing with motion planning. First collaboration was with the University of Utrecht in the Netherlands. We have proposed a method to evaluate and compare various environment decomposition techniques [74]. Environment decomposition is an important step to perform navigation planning in large static environments. Second collaboration was with the University of North Carolina in Chapel Hill (see Section 9.4.1.1). We have coupled a contact planner for virtual characters with ITOMP, a motion optimization technique to achieve complex motion in cluttered environment [69].

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Grants with Industry

8.1.1. Technicolor

Participants: Salma Jiddi, Eric Marchand.

no Univ. Rennes 1 15CC310-02D, duration: 36 months.

This project funded by Technicolor started in October 2015. It supports Salma Jiddi's Ph.D. about augmented reality (see Section 7.1.7).

8.1.2. Realyz

Participant: Eric Marchand.

no Inria Rennes 10822, duration: 36 months.

This project funded by Realiz started in October 2015. It is realized in cooperation with Anatole Lecuyer, Hybrid group at Irisa and Inria Rennes-Bretagne Atlantique to support Guillaume Cortes Ph.D. about motion capture.

8.1.3. Pôle Saint Hélier

Participants: Louise Devigne, Marie Babel.

no. Insa Rennes 2015/0890, duration: 36 months.

This project started in November 2015. It addresses the following two issues. First, the idea is to design a low-cost indoor / outdoor efficient obstacle avoidance system that respects the user intention, and does not alter user perception. This involves embedding innovative sensors to tackle the outdoor wheelchair navigation problem. The second objective is to take advantage of the proposed assistive tool to enhance the user Quality of Experience by means of biofeedback as well as the understanding of the evolution of the pathology.

8.1.4. Axyn

Participants: Dayana Hassan, Paolo Salaris, Patrick Rives.

no Inria Sophia 10874-1, duration: 36 months.

The objective of this project that started in November 2016 is to explore new methodologies for the interaction between humans and robots, autonomous navigation and mapping and to transfer the results obtained on the robotic platform developed by AXYN for assisting disabled/elderly people at home or in hospital structures. Cost limits, good accessibility to aged people, robustness and safety related to the applications are at the heart of the project. This contract (ANRT-CIFRE) support Dayana Hassan's Ph.D.

9. Partnerships and Cooperations

9.1. Regional Initiatives

9.1.1. ARED NavRob

Participants: Suman Raj Bista, Paolo Robuffo Giordano, François Chaumette.

no Inria Rennes 8033, duration: 36 months.

This project funded by the Brittany council ended in October 2016. It supported in part Suman Raj Bista's Ph.D. about visual navigation (see Section 7.4.1).

9.1.2. ARED DeSweep

Participants: Lesley-Ann Duflot, Alexandre Krupa.

no Inria Rennes 8033, duration: 36 months.

This project funded by the Brittany council started in October 2014. It supports in part Lesley-Ann Duflot's Ph.D. about visual servoing based on the shearlet transform (see Section 7.3.3).

9.1.3. ARED Locoflot

Participants: Ide Flore Kenmogne Fokam, Vincent Drevelle, Eric Marchand.

no Inria Rennes 9944, duration: 36 months.

This project funded by the Brittany council started in October 2015. It supports in part Ide Flore Kenmogne Fokam's Ph.D. about cooperative localization in multi-robot fleets using interval analysis (see Section 7.5.3).

9.1.4. ARED Mod4Nav

Participants: Aline Baudry, Marie Babel.

no INSA Rennes 2016/01, duration: 36 months.

This project funded by the Brittany council started in October 2016. It supports in part Aline Baudry's Ph.D. about wheelchair modeling.

9.1.5. "Equipement mi-lourd Rennes Métropole"

Participant: Paolo Robuffo Giordano.

no CNRS Rennes 14C0481, duration: 36 months.

A grant from "Rennes Métropole" has been obtained in June 2014 and supports the activities related to the use of drones (quadrotor UAVs). The platform described in Section 6.9.5 has been purchased in part thanks to this grant.

9.1.6. IRT Jules Verne Mascot

Participant: François Chaumette.

no Inria Rennes 10361, duration: 36 months.

This project started in October 2015. It is managed by IRT Jules Verne in Nantes and realized in cooperation with IRCCyN, Airbus, Renault, Faurecia and Alsthom. Its goal is to perform screwing for various industrial applications.

9.1.7. IRT b<>com NeedleWare

Participants: Hadrien Gurnel, Alexandre Krupa.

no Inria Rennes 9072, duration: 36 months.

This project started in October 2016. It supports Hadrien Gurnel's Ph.D. about the study of a shared control strategy fusing haptic and ultrasound visual control for assisting manual steering of needles for biopsies or therapy purposes in a synergetic way.

9.2. National Initiatives

9.2.1. France Life Imaging WP3-FLI ANFEET

Participant: Alexandre Krupa.

duration: 24 months.

This project started in January 2016. Its objective is to initiate collaborative research with the ICube laboratory (Strasbourg) on the control and supervision of flexible endoscopes in the digestive tube using ultrasound images.

9.2.2. ANR Contint Visioland

Participants: Noël Mériaux, Patrick Rives, François Chaumette.

no Inria Rennes 8304, duration: 48 months.

This project started in November 2013. It is composed of a consortium managed by Onera in Toulouse with Airbus, Spikenet Technology, IRCCyN, and Lagadic. Its aim is to develop vision-based localization and navigation techniques for autonomous landing on a runway (see Section 7.1.3).

9.2.3. ANR Contint Entracte

Participant: Julien Pettré.

no Inria Rennes 8013, duration: 42 months.

This project started in November 2013. It is realized in collaboration with the Gepetto group at Laas, Toulouse, and the Mimetic group at Irisa and Inria Rennes Bretagne Atlantique. It addresses the problem of motion planning for anthropomorphic systems, and more generally, the problem of manipulation path planning. ENTRACTE proposes to study in parallel both the mathematical foundation of artificial motion and the neurocognitive structures used by humans to quickly solve motion problems.

9.2.4. ANR JCJC Percolation

Participant: Julien Pettré.

no Inria Rennes 7991, duration: 42 months.

The ANR "Jeune Chercheur" Percolation project started on January 2014. It aims at designing perception-based crowd simulation algorithms. We develop agents which are capable of perceiving their virtual environment through virtual sensors, and which are able to navigate in it, as well as to interact with the other agents.

9.2.5. ANR JCJC SenseFly

Participants: Thomas Bellavoir, Muhammad Usman, Riccardo Spica, Paolo Robuffo Giordano.

no Irisa CNRS 50476, duration: 36 months.

The ANR "Jeune Chercheur" SenseFly project started in August 2015. Its goal is to advance the state-of-the-art in multi-UAV in the design and implementation of fully decentralized and sensor-based group behaviors by only resorting to onboard sensing (mainly cameras and IMU) and local communication (e.g., Bluetooth communication, wireless networks). Topics such as individual flight control, formation control robust against sensor limitations (e.g., limited field of view, occlusions), distributed estimation of relative positions/bearings from local sensing, maintenance of architectural properties of a multi-UAV formation will be touched by the project. Part of the platforms described in Section 6.9.5 has been purchased thanks to this grant.

9.2.6. ANR PLaTINUM

Participants: Eduardo Fernandez Moral, Vincent Drevelle, Patrick Rives.

no Inria Sophia 10204, duration: 42 months.

This project started in November 2015. It is composed of a consortium managed by Litis in Rouen with IGN Matis (Paris), Le2i (Le Creusot) and Lagadic group. It aims at proposing novel solutions to robust long-term mapping of urban environments.

9.2.7. BPI Romeo 2

Participants: Giovanni Claudio, Nicolas Cazy, Suman Raj Bista, Fabien Spindler, François Chaumette.

no Inria Rennes 7114, duration: 60 months.

This project started in November 2012. It is composed of a large consortium managed by Softbank Robotics (ex Aldebaran Robotics) with Laas in Toulouse, Isir in Paris, Lirmm in Montpellier, Inria groups Lagadic, Bipop (Pierre-Brice Wieber), Flowers (Pierre-Yves Oudeyer), and many other partners. It aims at developing advanced control and perception functionalities to a humanoid robot. In this project, we are in charge of visual manipulation and navigation with Romeo and Pepper. It supports in part Suman Raj Bista's Ph.D. about visual navigation (see Section 7.4.1), as well as Nicolas Cazy's Ph.D. about model-based predictive control for visual servoing (see Section 7.2.4).

9.2.8. Equipex Robotex

Participants: Fabien Spindler, François Chaumette.

no Inria Rennes 6388, duration: 9 years.

Lagadic is one of the 15 French academic partners involved in the Equipex Robotex network thats started in February 2011. It is devoted to get and manage significative equipment in the main robotics labs in France. In the scope of this project, we have got the humanoid robot Romeo (see Section 6.9.4).

9.3. European Initiatives

9.3.1. FP7 & H2020 Projects

9.3.1.1. FP7 Space RemoveDEBRIS

Participants: Aurélien Yol, Eric Marchand, François Chaumette.

Instrument: Specific Targeted Research Project

Duration: October 2013 - September 2017

Coordinator: University of Surrey (United Kingdom)

Partners: Surrey Satellite Technology (United Kingdom), Airbus (Toulouse, France and Bremen, Germany), Isis (Delft, The Netherlands), CSEM (Neuchâtel, Switzerland), Stellenbosch University (South Africa).

Inria contact: François Chaumette

Abstract: The goal of this project is to validate model-based tracking algorithms on images acquired during an actual space debris removal mission (see Section 7.1.2).

9.3.1.2. H2020 Comanoid

Participants: Don Joven Agravante, Giovanni Claudio, Souriya Trinh, Fabien Spindler, François Chaumette.

Title: Multi-contact Collaborative Humanoids in Aircraft Manufacturing

Programm: H2020

Duration: January 2015 - December 2018

Coordinator: CNRS (Lirmm)

Partners: Airbus Group (France), DLR (Germany), Università Degli Studi di Roma La Sapienza (Italy), CNRS (I3S)

Inria contact: Francois Chaumette

COMANOID investigates the deployment of robotic solutions in well-identified Airbus airliner assembly operations that are laborious or tedious for human workers and for which access is impossible for wheeled or rail-ported robotic platforms. As a solution to these constraints a humanoid robot is proposed to achieve the described tasks in real-use cases provided by Airbus Group. At a first glance, a humanoid robotic solution appears extremely risky, since the operations to be conducted are in highly constrained aircraft cavities with non-uniform (cargo) structures. Furthermore, these tight spaces are to be shared with human workers. Recent developments, however, in multi-contact planning and control suggest that this is a much more plausible solution than current alternatives such as a manipulator mounted on multi-legged base. Indeed, if humanoid robots can efficiently exploit their surroundings in order to support themselves during motion and manipulation, they can ensure balance and stability, move in non-gaited (acyclic) ways through narrow passages, and also increase operational forces by creating closed-kinematic chains. Bipedal robots are well suited to narrow environments specifically because they are able to perform manipulation using only small support areas. Moreover, the stability benefits of multi-legged robots that have larger support areas are largely lost when the manipulator must be brought close, or even beyond, the support borders. COMANOID aims at assessing clearly how far the state-of-the-art stands from such novel technologies. In particular the project focuses on implementing a real-world humanoid robotics solution using the best of research and innovation. The main challenge will be to integrate current scientific and technological advances including multi-contact planning and control; advanced visual-haptic servoing; perception and localization; human-robot safety and the operational efficiency of cobotics solutions in airliner manufacturing.

9.3.1.3. H2020 Romans

Participants: Nicolò Pedemonte, Firas Abi Farraj, Fabien Spindler, François Chaumette, Paolo Robuffo Giordano.

Title: Robotic Manipulation for Nuclear Sort and Segregation

Programm: H2020

Duration: May 2015 - April 2018

Coordinator: University of Birmingham

Partners: NLL (UK), CEA (France), Univ. Darmstadt (Germany)

CNRS contact: Paolo Robuffo Giordano

The RoMaNS (Robotic Manipulation for Nuclear Sort and Segregation) project will advance the state of the art in mixed autonomy for tele-manipulation, to solve a challenging and safety-critical "sort and segregate" industrial problem, driven by urgent market and societal needs. Cleaning up the past half century of nuclear waste, in the UK alone (mostly at the Sellafield site), represents the largest environmental remediation project in the whole of Europe. Most EU countries face related challenges. Nuclear waste must be "sorted and segregated", so that low-level waste is placed in low-level storage containers, rather than occupying extremely expensive and resource intensive high-level storage containers and facilities. Many older nuclear sites (>60 years in UK) contain large numbers of legacy storage containers, some of which have contents of mixed contamination levels, and sometimes unknown contents. Several million of these legacy waste containers must now be cut open, investigated, and their contents sorted. This can only be done remotely using robots, because of the high levels of radioactive material. Current state-of-the-art practice in the industry, consists of simple tele-operation (e.g. by joystick or teach-pendant). Such an approach is not viable in the long-term, because it is prohibitively slow for processing the vast quantity of material required. The project will: 1) Develop novel hardware and software solutions for advanced bi-lateral master-slave tele-operation. 2) Develop advanced autonomy methods for highly adaptive automatic grasping and manipulation actions. 3) Combine autonomy and tele-operation methods using state-of-the-art understanding of mixed initiative planning, variable autonomy and shared control approaches. 4) Deliver a TRL 6 demonstration in an industrial plant-representative environment at the UK National Nuclear Lab Workington test facility.

9.3.2. Collaborations with European Partners

Participants: Fabien Spindler, Alexandre Krupa, François Chaumette.

Project acronym: i-Process

Project title: Innovative and Flexible Food Processing Technology in Norway

Duration: January 2016 - December 2019

Coordinator: Sintef (Norway)

Other partners: Nofima, Univ. of Stavanger, NMBU, NTNU (Norway), DTU (Denmark), KU Leuven (Belgium), and about 10 Norwegian companies.

Abstract: This project is granted by the Norwegian Government. Its main objective is to develop novel concepts and methods for flexible and sustainable food processing in Norway. In the scope of this project, the Lagadic group is involved for visual tracking and visual servoing of generic and potentially deformable objects. Prof. Pal Johan from the Norwegian University of Life Sciences (NMBU), and Ekrem Misimi from Sintef spent a short visit in June and October respectively.

9.4. International Initiatives

9.4.1. Inria Associate Teams

9.4.1.1. SIMS

Title: Realistic and Efficient Simulation of Complex Systems

International Partners:

University of North Carolina at Chapel Hill (USA) - GAMMA Group - Ming C. Lin, Dinesh Manocha

University of Minnesota (USA) - Motion Lab - Stephen Guy

Brown University (USA) - VenLab - William Warren

Start year: 2012

See http://people.rennes.inria.fr/Julien.Pettre/EASIMS/easims.html

The general goal of SIMS is to make significant progress toward realistic and efficient simulation of highly complex systems, which raise combinatory explosive problems. This proposal is focused on human motion and interaction, and covers 3 active topics with wide application range:

1. Crowd simulation: virtual human interacting with other virtual humans,
2. Autonomous virtual humans interacting with their environment,
3. Physical simulation: real humans interacting with virtual environments.

SIMS is orthogonally structured by transversal questions: the evaluation of the level of realism reached by a simulation (which is a problem by itself in the considered topics), considering complex systems at various scales (micro, meso and macroscopic ones), and facing combinatory explosion of simulation algorithms.

9.4.1.2. ISI4NAVE

Title: Innovative Sensors and adapted Interfaces for assistive NAVigation and pathology Evaluation

International Partner:

University College London (United Kingdom) - Aspire CREATe - Tom Carlson

Start year: 2016

See http://www.irisa.fr/lagadic/team/MarieBabel/ISI4NAVE/ISI4NAVE.html

The global ageing population, along with disability compensation constitute major challenging societal and economic issues. In particular, achieving autonomy remains a fundamental need that contributes to the individual's wellness and well-being. In this context, innovative and smart technologies are designed to achieve independence while matching user's individual needs and desires.

Hence, designing a robotic assistive solution related to wheelchair navigation remains of major importance as soon as it compensates partial incapacities. This project will then address the following two issues. First, the idea is to design an indoor / outdoor efficient obstacle avoidance system that respects the user intention, and does not alter user perception. This involves embedding innovative sensors to tackle the outdoor wheelchair navigation problem. The second objective is to take advantage of the proposed assistive tool to enhance the user Quality of Experience by means of biofeedback. Indeed, adapted interfaces should improve the understanding of people that suffer from cognitive and/or visual impairments.

The originality of the project is to continuously integrate medical validation as well as clinical trials during the scientific research work in order to match user needs and acceptation.

9.4.2. Inria International Partners

9.4.2.1. Informal International Partners

- Alexandre Krupa has a collaboration with Prof. Nassir Navab from the Technical University of Munich concerning the joint supervision of Pierre Chatelain's Ph.D. (see Section 7.3.2).

9.4.3. Participation in International Programs

The Lagadic group is one of the few external partners of the Australian Center for Robotic Vision (see http://roboticvision.org). It groups QUT in Brisbane, ANU in Canberra, Monash University and Adelaide University. In the scope of this project, Peter Corke and Ben Upcroft spent a short visit in May 2016 while Jurgen Leitner spent a 1-month visit in October 2016.

9.5. International Research Visitors

9.5.1. Visits of International Scientists

- Nicolas Alt, senior researcher at Technical University of Munich (TUM) was a visiting scientist at Sophia Antipolis from Jan until Feb 2016. He worked on visuo-haptic environment perception.
- Alejandro Perez Yus, Ph.D. student at Universidad de Zaragoza, spent a 3-month visit in Sophia Antipolis from Sep until Nov 2016. He worked on the calibration of multi-camera RGB-D systems.
- Prof. Denis Wolf, associate professor at Univ. Sao Paulo, Brasil, spends a sabbatical year in Sophia Antipolis from Jul 2016 to Aug 2017. He works on semantic learning applied to intelligent vehicles.
- Nicola Battilani, Ph.D. student at University of Modena and Reggio Emilia, spent a 6-month visit in Rennes from May until Oct 2016. He worked on shared control algorithms for optimal 3D reconstruction from vision.

- Prof. Volkan Isler from University of Minnesota, Phillip Schmidt, Ph.D. student from DLR, Prof. Ivan Petrovic from Univ. of Zagreb, Prof. Purang Abolmaesumi from Univ. of British Columbia, Prof. Nassir Navab from Technical University of Munich, and Prof. Russ Taylor from John Hopkins University spent a short visit in the group in 2016.

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events Organization

10.1.1.1. General Chair, Scientific Chair

- Marie Babel was the Scientific Chair of the workshop "Innovation Robotique et Santé : Assistance à la conduite de fauteuil roulant" organized in Inria Rennes on December 2016 (150 participants).

10.1.1.2. Member of the Organizing Committees

- François Chaumette was in charge of organizing a Tutorial on Vision for Robotics at ICRA'2016: http://www.icra2016.org/conference/tutorials. The plenary speakers were Peter Corke (QUT, Australia), Jana Kosecka (George Mason University, US), Eric Marchand [45] and François Chaumette [44]. Around 200 participants attended this tutorial.
- Patrick Rives was in charge of co-organizing the "Journée Transports Intelligents " on behalf of the RFIA'2016 conference in Clermont-Ferrand: http://rfia2016.iut-auvergne.com/index.php/rfp-ia
- Paolo Robuffo Giordano has co-organized the Invited Session "Rigidity Theory for Problems in Multi-Agent Coordination" at the 54$^{\text{th}}$ IEEE Conf. on Decision and Control (CDC 2015), together by D. Zelazo (Technion, Israel) and A. Franchi (LAAS, France).

10.1.2. Scientific Events Selection

10.1.2.1. Chair of Conference Program Committees

- Paolo Robuffo Giordano: Workshop/Tutorial Session Chair for IROS 2016, Daejeon, Korea

10.1.2.2. Member of the Conference Program Committees

- François Chaumette: ICRA'2016
- Eric Marchand: ICRA 2016
- Patrick Rives: ICRA'2016, CVPR'2016
- Paolo Robuffo Giordano: ICRA'2016

10.1.2.3. Reviewer

- Marie Babel: IROS'2016 (2), ICIP'2016 (4)
- François Chaumette: IROS'2016 (1), ICRA'2017 (1)
- Vincent Drevelle: IFAC 2017 World Congress (1).
- Alexandre Krupa: IROS'2016 (1), ICRA'2017 (1)
- Eric Marchand: IROS 2016, ICRA 2017 (1), RFIA 2016 (2)
- Julien Pettré: SIGGRAPH 2016 (3), Collective Dynamics (1), CASA 2016 (3), IROS 2016 (3), ALIFE 2016 (2), ACM SCA 2016 (1), ACM CIE (1), TVCG (1), SIGGRAPH ASIA 2016 (1), Elsevier CAG (1), Pacific Graphics 2016 (1), Plos ONE (1), ACM MIG 2016 (2), IEEE VR 2017 (1), IEEE TBME (1), Taylor and Francis JMB (1), Eurographics 2017 (4), Elsevier TRB (1).
- Patrick Rives: ICRA'2016, ITSC'2016, CVPR'2016, ECCV'2016, IV'2016, ICINCO'2016, RFIA'2016, ICRA'2017

- Paolo Robuffo Giordano: ACC'2017 (1), DARS'2016 (2), ICRA'2017(2), IROS'2016 (3), RSS'2016 (1)

10.1.3. Journal

10.1.3.1. Member of the Editorial Boards

- François Chaumette: Editorial Board of the Int. Journal of Robotics Research, Senior Editor of the IEEE Robotics and Automation Letters, Editorial Board of the Springer Tracts in Advanced Robotics, Board Member of the Springer Encyclopedia of Robotics.
- Alexandre Krupa: Associate Editor of the IEEE Robotics and Automation Letters
- Eric Marchand: Associate Editor of the IEEE Robotics and Automation Letters
- Julien Pettré: Associate Editor for Computer Animation and Virtual Worlds, Associate Editor for Collective Dynamics
- Paolo Robuffo Giordano: Associate Editor of the IEEE Transactions on Robotics

10.1.3.2. Reviewer - Reviewing Activities

- Marie Babel: IEEE Trans. on Human-Machine Systems (1)
- François Chaumette: IEEE Trans. on Robotics (2), IEEE/ASME Trans. on Mechatronics (1), Robotics and Autonomous Systems (1), Journal of Intelligent and Robotic Systems (1), Control Engineering Practice (1)
- Vincent Drevelle: Transportation Research Part C (1)
- Eric Marchand : Int. Journal of robotics Research (1), Software, Practice and Experience (1), IEEE Trans. on Visualization and Computer Graphics (1), Visual Computer (1)
- Patrick Rives: IEEE Robotics and Automation Letters (1), Robotics and Autonomous Systems (1)
- Paolo Robuffo Giordano: IEEE Robotics and Automation letters (2), IEEE Trans. on Control of Network Systems (1), IEEE Trans. on Haptics (1)

10.1.4. Invited Talks

- François Chaumette: Plenary talk at RCAR'2016, Angkor Wat, Cambodia [43].
- Paolo Robuffo Giordano. Invited Talk: Collective Control, State Estimation and Human Interaction for Quadrotors in Unstructured Environments. 2016 GIS Micro-Drones Day, ENAC, Toulouse, France, October 2016
- Paolo Robuffo Giordano. Invited Seminar: Estimation and Control for Multi-Robot Systems. 2016 IEEE RAS Summer School on Multi-Robot Systems, Singapore, June 2016

10.1.5. Leadership within the Scientific Community

- François Chaumette is a 2016-2019 elected member of the Administrative Committee of the IEEE Robotics and Automation Society. He is also a member of the Scientific Council of the CNRS INS2I.
- François Chaumette and Patrick Rives are members of the scientific council of the "GdR Robotique".

10.1.6. Scientific Expertise

- François Chaumette is vice-president of the ANR Tremplin-ERC program (in charge of providing grants to recipients of non-funded A and B Starting and Consolidator ERC proposals). He was also a member of the "Institut Universitaire de France (IUF)" selection committee in charge of evaluating senior proposals, and served in the jury to select an Irstea senior researcher (DR2).
- Julien Pettré is project proposal reviewer for the Netherlands Organisation for Scientific Research.

- Paolo Robuffo Giordano is a reviewer for EU FP7 projects, for the ANR (French National Research Agency) ASTRID 2016 Program, for the “Comité ECOS Nord” of the “Ministère des affaires étrangères et du développement international” and the “Ministère de l’Enseignement supérieur et de la Recherche”, and for the DGA as expert on algorithms and sensors for drones.

10.1.7. Research Administration

- François Chaumette serves as the president of the committee in charge of all the temporary recruitments (“Commission Personnel”) at Inria Rennes-Bretagne Atlantique and Irisa. He is also a member of the Head team of Inria Rennes-Bretagne Atlantique.
- Alexandre Krupa and Julien Pettré are members of the CUMIR (“Commission des Utilisateurs des Moyens Informatiques pour la Recherche”) of Inria Rennes-Bretagne Atlantique.
- Eric Marchand served as secretary in the board of the “Association Française pour la Reconnaissance et l’Interprétation des Formes (AFRIF)”. He is also in charge of the Irisa Ph.D. students in the committee in charge of all the temporary recruitments (“Commission Personnel”) at Inria Rennes-Bretagne Atlantique and Irisa. He is in the board of the “Pôle Images et Réseaux” and in the board of “Ecole doctorale Matisse”.
- Julien Pettré is an elected member of the “Comité de Centre” at Inria Rennes-Bretagne Atlantique.
- Patrick Rives is a member of the “Comité des projets” and “Comité de Centre” at Inria Sophia Antipolis-Méditerranée.

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

Marie Babel:

Master INSA2: “Robotics”, 26 hours, M1, INSA Rennes

Master INSA1: “Architecture”, 30 hours, L3, INSA Rennes

Master INSA2: “Computer science project”, 30 hours, M1, INSA Rennes

Master INSA2: “Image analysis”, 18 hours, M1, INSA Rennes

Master INSA1: “Remedial math courses”, 50 hours, L3, INSA Rennes

François Chaumette:

Master ESIR3: “Visual servoing”, 8 hours, M2, Ecole supérieure d’ingénieurs de Rennes

Vincent Drevelle:

Master ESIR2: “Real-time systems and RTOS”, 24 hours, M1, Esir Rennes

Master GLA: “Terrain information systems”, 14 hours, M2, Université de Rennes 1

Master Info: “Artificial intelligence”, 12 hours, M1, Université de Rennes 1

Licence Info: “Computer systems architecture”, 20 hours, L1, Université de Rennes 1

Licence Miage: “Computer programming”, 78 hours, M1, Université de Rennes 1

Master CTS: “Instrumentation, localization, GPS”, 4 hours, M2, Université de Rennes 1

Licence and Master ET: “Electronics project”, 23 hours, L3 and M1, Université de Rennes 1

Alexandre Krupa:

Master SIBM (Signals and Images in Biology and Medicine): “Medical robotics guided from images”, 4.5 hours, M2, Université de Rennes 1, Brest and Angers

Master FIP TIC-Santé: “Ultrasound visual servoing”, 6 hours, M2, Télécom Physique Strasbourg

Master INSA3: “Modeling and engineering for Biology and Health applications”, 12 hours, M2, INSA Rennes

Master ESIR3: “Ultrasound visual servoing”, 9 hours, M2, Esir Rennes

Eric Marchand:

Master Esir2: “Colorimetry”, 24 hours, M1, Esir Rennes

Master Esir2: “Computer vision: geometry”, 24 hours, M1, Esir Rennes

Master Esir3: “Special effects”, 24 hours, M2, Esir Rennes

Master Esir3: “Computer vision: tracking and recognition”, 24 hours, M2, Esir Rennes

Master MRI: “Computer vision”, 24 hours, M2, Université de Rennes 1

Master MIA: “Augmented reality”, 4 hours, M2, Université de Rennes 1

Julien Pettré:

Licence Info : “Programmation Informatique”, 40, LI1, Université de Rennes 1, Rennes

INSA1: “Programmation Informatique”, 40 hours, INSA Rennes

10.2.2. Supervision

- Ph.D.: Le Cui, “Robust micro/nano-positioning by visual servoing”, defended on January 2016, supervised by Eric Marchand [15]
- Ph.D.: Vishnu Karakkat Narayanan, “Semi-autonomous navigation of a wheelchair by visual servoing and user intention analysis”, defended in November 2016, supervised by Marie Babel and Anne Spalanzani (Chroma group at Inria Rhône-Alpes) [16]
- Ph.D.: Nicolas Cazy, “Commande prédictive pour la réalisation de tâches d’asservissement visuel successives”, defended in November 2016, supervised by Paolo Robuffo Giordano, François Chaumette and Pierre-Brice Wieber (Bipop group at Inria Rhône-Alpes) [13]
- Ph.D.: Julien Bruneau, “Studying and modeling complex interactions for crowd simulation”, defended in November 2016, supervised by Julien Pettré and Anne-Hélène Olivier (Mimetic group at Inria Rennes-Bretagne Atlantique and Irisa) [12]
- Ph.D.: Aly Magassouba, “Aural servo: towards an alternative approach to sound localization for robot motion control”, defended in December 2016, supervised by François Chaumette and Nancy Bertin (Panama group at Inria Rennes-Bretagne Atlantique and Irisa)
- Ph.D.: Lucas Royer, “Real-time tracking of deformable targets in 3D ultrasound sequences”, defended in December 2016, supervised by Alexandre Krupa, Maud Marchal (Hybrid group at Inria Rennes-Bretagne Atlantique and Irisa) and Guillaume Dardenne (IRT b<>com)
- Ph.D.: Pierre Chatelain, “Quality-driven control of a robotized ultrasound probe”, defended in December 2016, supervised by Alexandre Krupa and Nassir Navab (Technische Universität München) [14]
- Ph.D.: Suman Raj Bista, “Indoor Navigation of Mobile Robots based on Visual Memory and Image-Based Visual Servoing”, defended in December 2016, supervised by Paolo Robuffo Giordano and François Chaumette [11]
- Ph.D. in progress: Renato José Martins, “Robust navigation and control of an autonomous vehicle”, started in November 2013, supervised by Patrick Rives and Samuel Siqueira Bueno (CTI)
- Ph.D. in progress: Jason Chevrie, “Control of a flexible needle by visual servoing using B-mode ultrasound images”, started in September 2014, supervised by Alexandre Krupa and Marie Babel
- Ph.D. in progress: Quentin Bateux, “Visual servoing from global descriptors”, started in October 2014, supervised by Eric Marchand
- Ph.D. in progress: Fabrizio Schiano, “Collective control with onboard sensors for multiple quadrotor UAVs”, started in October 2014, supervised by Paolo Robuffo Giordano
- Ph.D. in progress: Pedro Patlan-Rosales, “Enhancement of ultrasound elastography by visual servoing and force control”, started in October 2014, supervised by Alexandre Krupa

- Ph.D. in progress: Noël Mériaux, “Landing by visual servoing”, started in October 2014, supervised by François Chaumette, Eric Marchand and Patrick Rives
- Ph.D. in progress: Lesley-Ann Duflot, “Visual servoing using shearlet transform”, started in November 2014, supervised by Alexandre Krupa and Brahim Tamadazte (Minarob group at FEMTO-ST, Besançon)
- Ph.D. in progress: Firas Abi Farraj, “Shared Control Architectures for Visual Servoing Tasks”, started in October 2015, supervised by Paolo Robuffo Giordano
- Ph.D. in progress: Salma Jiddi, “Analyses géométrique et photométrique pour des applications de réalité mixte”, started in October 2015, supervised by Eric Marchand and Philippe Robert (Technicolor)
- Ph.D. in progress: Ide Flore Kenmogne Fokam, “Cooperative localization in multi-robot fleets using interval analysis”, started in October 2015, supervised Vincent Drevelle and Eric Marchand
- Ph.D. in progress: Bryan Penin “Model predictive visual servoing for UAVS”, started in October 2015, supervised by Paolo Robuffo Giordano and François Chaumette
- Ph.D. in progress: Guillaume Cortes, “Motion Capture”, started in October 2015, supervised Eric Marchand and Anatole Lecuyer.
- Ph.D. in progress: Muhammad Usman, “Robust Vision-Based Navigation for Quadrotor UAVs”, started in October 2015, supervised by Paolo Robuffo Giordano
- Ph.D. in progress: Louise Devigne, “Contribution d’une aide technique robotique à l’évaluation de pathologies neurologiques : Application à la navigation d’un fauteuil roulant”, started in November 2015, supervised by Marie Babel and Philippe Gallien (Pôle Saint Hélier)
- Ph.D. in progress: Quentin Delamare, “Algorithmes d’estimation et de commande pour des quadrirotors en interaction physique avec l’environnement”, started in September 2016, supervised by Paolo Robuffo Giordano
- Ph.D. in progress: Axel Lopes, “Data assimilation for synthetic vision-based crowd simulation algorithms”, started in October 2016, supervised by Julien Pettré and François Chaumette
- Ph.D. in progress: Aline Baudry, “Contribution à la modélisation des fauteuils roulants pour l’amélioration de leur navigation en mode semi-autonome”, started in October 2016, supervised by Marie Babel and Sylvain Guégan (Mechanical Engineering Dpt/LGCGM at Insa Rennes)
- Ph.D. in progress: Hadrien Gurnel, “Shared control of a biopsie needle from haptic and ultrasound visual feedback”, started in October 2016, supervised by Alexandre Krupa and Maud Marchal (Hybrid group at Inria Rennes-Bretagne Atlantique and Irisa)
- Ph.D. in progress: Dayana Hassan, “Plate-forme robotisée d’assistance aux personnes à mobilité réduite”, started in November 2016, supervised by Paolo Salaris, Patrick Rives and Frank Anjeaux (Axyn robotique)
- Internship: Valentin Bureau from May 2016 until Sep 2016 (2 months in mobility in University College of London within ISI4NAVE associated team), M1/Second year of Computer Science department, INSA Rennes, supervised by Marie Babel
- Internship: Timothée Collard “Adapting a localization application to ROS and MAV navigation” from Jun 2016 until Aug 2016, M1, Université de Rennes 1, supervised by Vincent Drevelle
- Internship: Benjamin Fasquelle from May 2016 until Aug 2016, M1, ENS Rennes, supervised by Eric Marchand
- Benoit Heintz from Mar 2016 until Jul 2016, L2 level, ENSIL Limoges, supervised by Fabien Spindler and Giovanni Claudio
- Internship: Manutea Huang from Jun 2016 until Sep 2016, L3/First year of Computer Science department, INSA Rennes, supervised by Marie Babel
- Internship: Daniel Huc from May 2016 until Sep 2016, M2, IM2AG Grenoble, supervised by Julien Pettré

- Internship: Valentin Limantour from Sep 2016 until December 2016, M1, ENIB Brest, supervised by Paolo Robuffo Giordano
- Internship: Etienne Moisdon from May 2016 until Jun 2016, M, Univ. Rennes 1, supervised by Julien Pettré

10.2.3. External Ph.D. and HdR Juries

- François Chaumette: Julien Bruneau (Ph.D., president, Inria Rennes), Andrea Cherubini (HdR, president, Lirmm, Montpellier)
- Alexandre Krupa: Laure-Anais Chanel (Ph.D., reviewer, ICube, Strasbourg), Mouloud Ourak (Ph.D., reviewer, FEMTO-ST, Besançon), Paul Mignon (Ph.D., reviewer, TIMC-IMAG, Grenoble)
- Eric Marchand: Vishnu Karakkat Narayanan (Ph.D., president, Inria Rennes), Suman Raj Bista (Ph.D., president, Inria Rennes), Limming Yang (Ph.D., reviewer, EC Nantes), Tu-Hoa Pham (Ph.D., reviewer, Lirmm, Montpellier).
- Julien Pettré: Christian Vassallo (Ph.D., Laas, Toulouse), Fabien Cissé (Ph.D., UPMC, Paris), Thomas Pitiot (PH.D., ICube, Strasbourg)
- Patrick Rives: Zui Tao (Ph.D., president, UTC, Compiègne), Bruno Ricaud (Ph.D., reviewer, Ecole des Mines, Paris), Victor Gibert (Ph.D., member, IRCCyN, Nantes), Yue Kang (Ph.D., reviewer, UTC, Compiègne), Abdelhamid Dine (Ph.D., president, ENS-Cachan, Paris-Saclay), Pierre Merriaux (Ph.D., reviewer, Esigelec, Rouen), Fabrice Mayran de Chamiso (Ph.D., member, CEA-LIST, Paris-Saclay), Bruno Vallet (HdR, member, IGN, Paris).
- Paolo Robuffo Giordano: Osamah Saif (Ph.D., UTC, Compiègne, France), Marco Aggravi (Ph.D., University of Siena, Italy), Leonardo Meli (Ph.D., University of Siena, Italy).

10.3. Popularization

- Due to the visibility of our experimental platforms, the team is often requested to present its research activities to students, researchers or industry. Our panel of demonstrations allows us to highlight recent results concerning the positioning of an ultrasound probe by visual servoing, grasping and dual arm manipulation by Romeo, vision-based shared control using our haptic device for object manipulation, the control of a fleet of quadrotors, vision-based detection and tracking for space navigation in a rendezvous context, the semi-autonomous navigation of a wheelchair, and augmented reality applications. Some of these demonstrations are available as videos on VispTeam YouTube channel (https://www.youtube.com/user/VispTeam/videos). This year there were among others, demonstrations organized for the HCERES expert committee that evaluated IRISA, about twenty people affiliated to the CNRS electronics network, students of the “Innovation et entreprenariat” Master, those of the L3 R&I at ENS Rennes, about twenty students from the “Ecole des Mines de Nancy”, several classes of high school students around Rennes, without forgetting the members of the Ph.D. thesis juries.
- Fabien Spindler and Giovanni Claudio were interviewed by TV Rennes about Pepper robot (https://www.facebook.com/166100743494694/videos/vb.166100743494694/959489554155805/?type=2&theater).
- Marie Babel participated to “15^{e} Journée nationale des pôles de compétitivité” in March 2016 (Paris): HandiViz project was selected by the French Ministry of Finance
- Marie Babel participated to the Science Festival in October 2016 with an interview at “Village des Sciences” and a workshop for general public organized in Acigné near Rennes.
- Marie Babel participated to the “Convention Nationale des SATT” (October 2016, Paris).
- Marie Babel gave a talk on "HandiViz: a new driving experience" in October 2016, during the “Semaine des Technologies - Robotique et santé” organized at Insa Rennes.

- Vincent Drevelle participated to the “Journée science et musique” in Rennes, with an interactive demonstration of sound-based tracking of a micro aerial vehicle with a beam projector (in cooperation with the Panama team).
- An article related to the research activity of Alexandre Krupa on robotic needle steering entitled “Un robot qui apprend à viser” has been published in March 2016 in the general-audience magazine “Sciences Ouest”: http://www.espace-sciences.org/sciences-ouest/340/dossier/un-robot-qui-apprend-a-viser.
- Paolo Robuffo Giordano has given press releases on the activities involving formation control for multiple quadrotor UAVs to Émergences Inria, Sciences Ouest, and “Industrie & Technologies”.

11. Bibliography

Major publications by the team in recent years

[1] F. CHAUMETTE, S. HUTCHINSON.*Visual servoing and visual tracking*, in "Handbook of Robotics", B. SICILIANO, O. KHATIB (editors), Springer, 2008, chap. 24, p. 563-583, http://hal.inria.fr/hal-00920414/en.

[2] A. COMPORT, E. MARCHAND, M. PRESSIGOUT, F. CHAUMETTE.*Real-time markerless tracking for augmented reality: the virtual visual servoing framework*, in "IEEE Trans. on Visualization and Computer Graphics", July 2006, vol. 12, n^{o} 4, p. 615–628, https://hal.inria.fr/inria-00161250.

[3] A. DAME, E. MARCHAND.*Second order optimization of mutual information for real-time image registration*, in "IEEE Trans. on Image Processing", 2012, vol. 21, n^{o} 9, p. 4190-4203, http://hal.inria.fr/hal-00750528/en.

[4] A. DIOSI, S. SEGVIC, A. REMAZEILLES, F. CHAUMETTE.*Experimental Evaluation of Autonomous Driving Based on Visual Memory and Image Based Visual Servoing*, in "IEEE Trans. on Intelligent Transportation Systems", September 2011, vol. 12, n^{o} 3, p. 870–883, http://hal.inria.fr/hal-00639680/en.

[5] E. MARCHAND, F. SPINDLER, F. CHAUMETTE.*ViSP for visual servoing: a generic software platform with a wide class of robot control skills*, in "IEEE Robotics and Automation Magazine", December 2005, vol. 12, n^{o} 4, p. 40-52, https://hal.inria.fr/inria-00351899.

[6] R. MEBARKI, A. KRUPA, F. CHAUMETTE.*2D ultrasound probe complete guidance by visual servoing using image moments*, in "IEEE Trans. on Robotics", April 2010, vol. 26, n^{o} 2, p. 296-306, https://hal.inria.fr/inria-00544791.

[7] M. MEILLAND, A. COMPORT, P. RIVES.*Dense omnidirectional RGB-D mapping of large scale outdoor environments for real-time localisation and autonomous navigation*, in "Journal of Field Robotics, Special Issue on Ground Robots Operating in dynamic, unstructured and large-scale outdoor environments", June 2015, vol. 32, n^{o} 4, p. 474-503, http://hal.inria.fr/hal-01010429.

[8] C. NADEAU, A. KRUPA.*Intensity-based ultrasound visual servoing: modeling and validation with 2D and 3D probes*, in "IEEE. Trans. on Robotics", August 2013, vol. 29, n^{o} 4, p. 1003-1015 [*DOI :* 10.1109/TRO.2013.2256690], http://hal.inria.fr/hal-00854100.

[9] R. SPICA, P. ROBUFFO GIORDANO, F. CHAUMETTE.*Active Structure from Motion: Application to Point, Sphere and Cylinde*, in "IEEE Trans. on Robotics", December 2015, vol. 30, n^{o} 6, p. 1499-1513, http://hal.inria.fr/hal-01010429.

[10] D. ZELAZO, A. FRANCHI, H. H. BÜLTHOFF, P. ROBUFFO GIORDANO.*Decentralized rigidity maintenance control with range measurements for multi-robot systems*, in "The Int. Journal of Robotics Research", January 2015, vol. 34, n^o 1, p. 105-128, http://hal.inria.fr/hal-01076423.

Publications of the year

Doctoral Dissertations and Habilitation Theses

[11] S. R. BISTA.*Indoor Navigation of Mobile Robots based on Visual Memory and Image-Based Visual Servoing*, Universite de Rennes 1 ; Inria Rennes Bretagne Atlantique, December 2016, https://tel.archives-ouvertes.fr/tel-01426763.

[12] J. BRUNEAU.*Studying and modeling complex interactions for crowd simulation*, Université de Rennes 1 [UR1], November 2016, https://hal.inria.fr/tel-01425268.

[13] N. CAZY.*Predictive control for the achievement of successive visual servoing tasks*, Université Rennes 1, November 2016, https://tel.archives-ouvertes.fr/tel-01421363.

[14] P. CHATELAIN.*Quality-Driven Control of a Robotized Ultrasound Probe*, Université de Rennes 1 ; Technische Universität München, December 2016, https://tel.archives-ouvertes.fr/tel-01426511.

[15] L. CUI.*Robust micro/nano-positioning by visual servoing*, Université Rennes 1, January 2016, https://tel.archives-ouvertes.fr/tel-01267585.

[16] V. KARAKKAT NARAYANAN.*Characterizing assistive shared control through vision-based and human-aware designs for wheelchair mobility assistance* , Inria Rennes - Bretagne Atlantique ; INSA Rennes, November 2016, https://tel.archives-ouvertes.fr/tel-01426748.

[17] A. MAGASSOUBA.*Aural servo: towards an alternative approach to sound localization for robot motion control*, Université Rennes 1, December 2016, https://tel.archives-ouvertes.fr/tel-01426710.

[18] P. ROBUFFO GIORDANO.*Contributions to shared control and coordination of single and multiple robots*, Université de Rennes 1, January 2016, Habilitation à diriger des recherches, https://tel.archives-ouvertes.fr/tel-01301644.

[19] L. ROYER.*Real-time Tracking of Deformable Targets in 3D Ultrasound Sequences*, INSA de Rennes, December 2016, https://tel.archives-ouvertes.fr/tel-01426711.

Articles in International Peer-Reviewed Journal

[20] D. J. AGRAVANTE, G. CLAUDIO, F. SPINDLER, F. CHAUMETTE.*Visual servoing in an optimization framework for the whole-body control of humanoid robots*, in "IEEE Robotics and Automation Letters", 2017, https://hal.inria.fr/hal-01421734.

[21] Q. BATEUX, É. MARCHAND.*Histograms-based Visual Servoing*, in "IEEE Robotics and Automation Letters", January 2017, vol. 2, n^o 1, p. 80-87, https://hal.inria.fr/hal-01265560.

[22] S. BRIOT, F. CHAUMETTE, P. MARTINET.*Revisiting the determination of the singularity cases in the visual servoing of image points through the concept of hidden robot*, in "IEEE Transactions on Robotics", January 2017, vol. 33, n^o 2, https://hal.archives-ouvertes.fr/hal-01399774.

[23] S. BRIOT, P. MARTINET, F. CHAUMETTE.*Determining the Singularities for the Observation of Three Image Lines*, in "IEEE Robotics and Automation Letters", April 2017, vol. 2, n^o 2, p. 412-419, https://hal.archives-ouvertes.fr/hal-01398925.

[24] J. BRUNEAU, J. PETTRÉ.*EACS: Effective Avoidance Combination Strategy*, in "Computer Graphics Forum", September 2016 [*DOI :* 10.1111/CGF.13066], https://hal.inria.fr/hal-01392248.

[25] V. DREVELLE, P. BONNIFAIT.*Interval-based fast fault detection and identification applied to radio-navigation multipath*, in "International Journal of Adaptive Control and Signal Processing", January 2016, vol. 30, p. 154–172 [*DOI :* 10.1002/ACS.2535], https://hal.inria.fr/hal-01258946.

[26] E. FERNÁNDEZ-MORAL, P. RIVES, V. ARÉVALO, J. GONZÁLEZ-JIMÉNEZ.*Scene structure registration for localization and mapping*, in "Robotics and Autonomous Systems", January 2016, vol. 75, n^o B, p. 649-660 [*DOI :* 10.1016/J.ROBOT.2015.09.009], https://hal.inria.fr/hal-01237845.

[27] J. FORSHAW, G. AGLIETTI, N. NAVARATHINAM, H. KADHEM, T. SALMON, A. PISSELOUP, E. JOFFRE, T. CHABOT, I. RETAT, R. AXTHELM, S. BARRACLOUGH, A. RATCLIFFE, C. BERNAL, F. CHAUMETTE, A. POLLINI, W. STEYN.*REMOVEDEBRIS: An In-Orbit Active Debris Removal Demonstration Mission*, in "Acta Astronautica", October 2016, vol. 127, p. 448-463 [*DOI :* 10.1016/J.ACTAASTRO.2016.06.018], https://hal.inria.fr/hal-01342268.

[28] A. FRANCHI, P. ROBUFFO GIORDANO.*Online Leader Selection for Improved Collective Tracking and Formation Maintenance*, in "IEEE Transactions on Control of Network Systems", 2016 [*DOI :* 10.1109/TCNS.2016.2567222], https://hal.archives-ouvertes.fr/hal-01315463.

[29] L. HOYET, A.-H. OLIVIER, R. KULPA, J. PETTRÉ.*Perceptual Effect of Shoulder Motions on Crowd Animations*, in "ACM Transactions on Graphics", July 2016 [*DOI :* 10.1145/2897824.2925931], https://hal.inria.fr/hal-01357713.

[30] V. KARAKKAT NARAYANAN, F. PASTEAU, M. MARCHAL, A. KRUPA, M. BABEL.*Vision-based adaptive assistance and haptic guidance for safe wheelchair corridor following*, in "Computer Vision and Image Understanding", August 2016, vol. 179, p. 171-185, https://hal.inria.fr/hal-01277585.

[31] A. KRUPA, D. FOLIO, C. NOVALES, P. VIEYRES, T. LI.*Robotized Tele-Echography: an Assisting Visibility Tool to Support Expert Diagnostic*, in "IEEE Systems Journal", September 2016, vol. 10, n^o 3, p. 974-983 [*DOI :* 10.1109/JSYST.2014.2314773], https://hal.archives-ouvertes.fr/hal-00986875.

[32] J. LIMA, F. SIMÕES, H. UCHIYAMA, V. TEICHRIEB, E. MARCHAND.*Depth-Assisted Rectification for Real-Time Object Detection and Pose Estimation*, in "Machine Vision and Applications", February 2016, vol. 27, n^o 2, p. 193-219, https://hal.inria.fr/hal-01233046.

[33] É. MARCHAND, H. UCHIYAMA, F. SPINDLER.*Pose Estimation for Augmented Reality: A Hands-On Survey*, in "IEEE Transactions on Visualization and Computer Graphics", December 2016, vol. 22, n^o 12, p. 2633 - 2651 [*DOI :* 10.1109/TVCG.2015.2513408], https://hal.inria.fr/hal-01246370.

[34] C. NADEAU, A. KRUPA, J. PETR, C. BARILLOT.*Moments-Based Ultrasound Visual Servoing: From Mono to Multi-plane Approach*, in "IEEE Transactions on Robotics", 2017 [*DOI :* 10.1109/TRO.2016.2604482], https://hal.inria.fr/hal-01385661.

[35] T. NESTMEYER, P. ROBUFFO GIORDANO, H. H. BÜLTHOFF, A. FRANCHI.*Decentralized Simultaneous Multi-target Exploration using a Connected Network of Multiple Robots*, in "Autonomous Robots", 2016 [*DOI :* 10.1007/S10514-016-9578-9], https://hal.inria.fr/hal-01332937.

[36] P. PAPADAKIS, P. RIVES.*Binding human spatial interactions with mapping for enhanced mobility in dynamic environments*, in "Autonomous Robots", 2016 [*DOI :* 10.1007/S10514-016-9581-1], https://hal.inria.fr/hal-01342255.

[37] F. PASTEAU, V. KARAKKAT NARAYANAN, M. BABEL, F. CHAUMETTE.*A visual servoing approach for autonomous corridor following and doorway passing in a wheelchair*, in "Robotics and Autonomous Systems", January 2016, vol. 75, part A, p. 28-40 [*DOI :* 10.1016/J.ROBOT.2014.10.017], https://hal.inria.fr/hal-01068163.

[38] S. RAJ BISTA, P. ROBUFFO GIORDANO, F. CHAUMETTE.*Appearance-based Indoor Navigation by IBVS using Line Segments*, in "IEEE Robotics and Automation Letters", January 2016, vol. 1, n^{o} 1, p. 423-430, Also presented in IEEE Int. Conf. on Robotics and Automation, Stockolm, Sweden, https://hal.inria.fr/hal-01259750.

[39] Z. REN, P. CHARALAMBOUS, J. BRUNEAU, Q. PENG, J. PETTRÉ.*Group Modeling: A Unified Velocity-Based Approach*, in "Computer Graphics Forum", 2016 [*DOI :* 10.1111/CGF.12993], https://hal.inria.fr/hal-01372766.

[40] L. ROYER, A. KRUPA, G. DARDENNE, A. LE BRAS, É. MARCHAND, M. MARCHAL.*Real-time Target Tracking of Soft Tissues in 3D Ultrasound Images Based on Robust Visual Information and Mechanical Simulation*, in "Medical Image Analysis", January 2017, vol. 35, p. 582 - 598 [*DOI :* 10.1016/J.MEDIA.2016.09.004], https://hal.inria.fr/hal-01374589.

[41] C. VASSALLO, A.-H. OLIVIER, P. SOUÈRES, A. CRÉTUAL, O. STASSE, J. PETTRÉ.*How do walkers avoid a mobile robot crossing their way?*, in "Gait and Posture", January 2017, vol. 51, p. 97-103, https://hal.archives-ouvertes.fr/hal-01371202.

[42] D. WOLINSKI, M. C. LIN, J. PETTRÉ.*WarpDriver: Context-Aware Probabilistic Motion Prediction for Crowd Simulation*, in "ACM Transactions on Graphics", November 2016, vol. 35, n^{o} 6 [*DOI :* 10.1145/2980179.2982442], https://hal.inria.fr/hal-01411087.

Invited Conferences

[43] F. CHAUMETTE.*Visual servoing with and without image processing*, in "Plenary talk, IEEE International Conference on Real-time Computing and Robotics, RCAR'2016", Angkor Wat, Cambodia, June 2016, https://hal.inria.fr/hal-01385400.

[44] F. CHAUMETTE.*Visual servoing*, in "Tutorial on Vision for Robotics, IEEE Int. Conf. on Robotics and Automation", Stockholm, Sweden, May 2016, https://hal.inria.fr/hal-01385401.

[45] E. MARCHAND.*Visual tracking*, in "Tutorial on Vision for Robotics, IEEE Int. Conf. on Robotics and Automation", Stockholm, Sweden, May 2016, https://hal.inria.fr/hal-01385405.

International Conferences with Proceedings

[46] F. ABI-FARRAJ, N. PEDEMONTE, P. ROBUFFO GIORDANO.*A Visual-Based Shared Control Architecture for Remote Telemanipulation*, in "IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, IROS'16", Daejeon, South Korea, October 2016, p. 4266-4273, https://hal.inria.fr/hal-01355785.

[47] G. ANDRADE, F. NOUVIALE, J. ARDOUIN, É. MARCHAND, M. MARCHAL, A. LÉCUYER.*Enjoy 360° Vision with the FlyVIZ*, in "SIGGRAPH 2016 Emerging Technologies", Anaheim, United States, July 2016 [*DOI :* 10.1145/2929464.2929471], https://hal.inria.fr/hal-01387573.

[48] Q. BATEUX, E. MARCHAND.*Particle Filter-based Direct Visual Servoing*, in "IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, IROS'16", Daejeon, South Korea, October 2016, p. 4180-4186, https://hal.inria.fr/hal-01355396.

[49] S. BRIOT, P. MARTINET, F. CHAUMETTE.*Singularity Cases in the Visual Servoing of Three Image Lines*, in "2017 IEEE International Conference on Robotics and Automation (ICRA 2017)", Singapour, Singapore, Proceedings of 2017 IEEE International Conference on Robotics and Automation (ICRA 2017), May 2017, https://hal.archives-ouvertes.fr/hal-01435811.

[50] P. CHATELAIN, A. KRUPA, N. NAVAB.*Confidence-Driven Control of an Ultrasound Probe: Target-Specific Acoustic Window Optimization*, in "IEEE Int. Conf. on Robotics and Automation, ICRA'16", Stockholm, Sweden, May 2016, https://hal.inria.fr/hal-01274748.

[51] J. CHEVRIE, A. KRUPA, M. BABEL.*Needle Steering Fusing Direct Base Manipulation and Tip-based Control*, in "IEEE Int. Conf. on Robotics and Automation, ICRA'16", Stockholm, Sweden, May 2016, https://hal.inria.fr/hal-01304860.

[52] J. CHEVRIE, A. KRUPA, M. BABEL.*Online prediction of needle shape deformation in moving soft tissues from visual feedback*, in "IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, IROS'16", Daejeon, South Korea, October 2016, p. 2357-2362, https://hal.inria.fr/hal-01355484.

[53] G. CLAUDIO, F. SPINDLER, F. CHAUMETTE.*Vision-based manipulation with the humanoid robot Romeo*, in "IEEE-RAS International Conference on Humanoid Robotics, Humanoids 2016", Cancun, Mexico, November 2016, https://hal.inria.fr/hal-01385408.

[54] L. CUI, E. MARCHAND, S. HALIYO, S. RÉGNIER.*Three-Dimensional Visual Tracking and Pose Estimation in Scanning Electron Microscopes*, in "IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, IROS'16", Daejeon, South Korea, October 2016, p. 5210-5215, https://hal.inria.fr/hal-01355393.

[55] L. DEVIGNE, V. KARAKKAT NARAYANAN, F. PASTEAU, M. BABEL.*Low complex sensor-based shared control for power wheelchair navigation*, in "IROS2016 - IEEE/RSJ International Conference on Intelligent Robots and Systems", Daejeon, South Korea, October 2016, p. 5434-5439, https://hal.inria.fr/hal-01355410.

[56] V. DREVELLE.*Convergence domain of image-based visual servoing with a line-scan camera*, in "Summer Workshop on Interval Methods, SWIM 2016", Lyon, France, June 2016, https://hal.inria.fr/hal-01415435.

[57] L.-A. DUFLOT, A. KRUPA, B. TAMADAZTE, N. ANDREFF.*Shearlet Transform: a Good Candidate for Compressed Sensing in Optical Coherence Tomography*, in "IEEE Int. Conf. on Engineering in Medicine and Biology Society, EMBC'16", Orlando, United States, August 2016, p. 435-438, https://hal.inria.fr/hal-01355488.

[58] L.-A. DUFLOT, A. KRUPA, B. TAMADAZTE, N. ANDREFF.*Shearlet-based vs. Photometric-based Visual Servoing for Robot-assisted Medical Applications*, in "IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, IROS'16", Daejeon, South Korea, October 2016, p. 4099-4104, https://hal.inria.fr/hal-01355414.

[59] L.-A. DUFLOT, A. KRUPA, B. TAMADAZTE, N. ANDREFF.*Toward Ultrasound-based Visual Servoing using Shearlet Coefficients*, in "IEEE Int. Conf. on Robotics and Automation, ICRA'16", Stockholm, Sweden, May 2016, https://hal.inria.fr/hal-01304753.

[60] S. JIDDI, P. ROBERT, E. MARCHAND.*Reflectance and Illumination Estimation for Realistic Augmentations of Real Scenes*, in "IEEE Int. Symp. on Mixed and Augmented Reality, ISMAR'16 (poster session)", Merida, Mexico, September 2016, Poster, https://hal.inria.fr/hal-01355581.

[61] V. KARAKKAT NARAYANAN, A. SPALANZANI, M. BABEL.*A semi-autonomous framework for human-aware and user intention driven wheelchair mobility assistance*, in "IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, IROS'16", Daejeon, South Korea, October 2016, p. 4700-4707, https://hal.inria.fr/hal-01355481.

[62] V. KARAKKAT NARAYANAN, A. SPALANZANI, R. C. LUO, M. BABEL.*Analysis of an adaptive strategy for equitably approaching and joining human interactions*, in "IEEE Int. Symp. on Robot and Human Interactive Communication, RO-MAN", New-York, United States, IEEE Int. Symp. on Robot and Human Interactive Communication, RO-MAN, August 2016, https://hal.inria.fr/hal-01330889.

[63] I.-F. KENMOGNE, V. DREVELLE.*Image-based Mobile Robot localization using Interval Methods*, in "Summer Workshop on Interval Methods, SWIM 2016", Lyon, France, June 2016, https://hal.inria.fr/hal-01415432.

[64] A. MAGASSOUBA, N. BERTIN, F. CHAUMETTE.*Audio-based robot controlfrom interchannel level difference and absolute sound energy*, in "IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, IROS'16", Daejeon, South Korea, October 2016, p. 1992-1999, https://hal.inria.fr/hal-01355394.

[65] A. MAGASSOUBA, N. BERTIN, F. CHAUMETTE.*Binaural auditory interaction without HRTF for humanoid robots: A sensor-based control approach*, in "Workshop on Multimodal Sensor-based Control for HRI and soft manipulation, IROS'2016", Daejeon, South Korea, October 2016, https://hal.inria.fr/hal-01408422.

[66] A. MAGASSOUBA, N. BERTIN, F. CHAUMETTE.*First applications of sound-based control on a mobile robot equipped with two microphones*, in "IEEE Int. Conf. on Robotics and Automation, ICRA'16", Stockholm, Sweden, May 2016, https://hal.inria.fr/hal-01277589.

[67] R. MARTINS, E. FERNANDEZ-MORAL, P. RIVES.*Adaptive Direct RGB-D Registration and Mapping for Large Motions*, in "Asian Conference on Computer Vision, ACCV 2016", Taipei, Taiwan, November 2016, https://hal.inria.fr/hal-01403953.

[68] R. MARTINS, P. RIVES.*Increasing the Convergence Domain of RGB-D Direct Registration Methods for Vision-based Localization in Large Scale Environments*, in "Workshop on Planning, Perception and Navigation for Intelligent Vehicles – IEEE Intelligent Transportation Systems Conference, ITSC PPNIV", Rio de Janeiro, Brazil, November 2016, https://hal.inria.fr/hal-01403961.

[69] C. PARK, J. S. PARK, S. TONNEAU, N. MANSARD, F. MULTON, J. PETTRÉ, D. MANOCHA.*Dynamically balanced and plausible trajectory planning for human-like characters*, in "20th ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games", Redmond, United States, ACM, February 2016, p. 39-48 [*DOI :* 10.1145/2856400.2856405], https://hal.inria.fr/hal-01290368.

[70] P. PATLAN-ROSALES, A. KRUPA.*Automatic palpation for quantitative ultrasound elastography by visual servoing and force control*, in "IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, IROS'16", Daejeon, South Korea, October 2016, p. 2357-2362, https://hal.inria.fr/hal-01355406.

[71] S. RAJ BISTA, P. ROBUFFO GIORDANO, F. CHAUMETTE.*Appearance-based Indoor Navigation by IBVS using Mutual Information*, in "IEEE Int. Conf. on Control, Automation, Robotics and Vision, ICARCV 2016", Phuket, Thailand, November 2016, https://hal.inria.fr/hal-01355382.

[72] F. SCHIANO, A. FRANCHI, D. ZELAZO, P. ROBUFFO GIORDANO.*A Rigidity-Based Decentralized Bearing Formation Controller for Groups of Quadrotor UAVs*, in "IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2016", Daejeon, South Korea, October 2016, p. 5099-5106, https://hal.archives-ouvertes.fr/hal-01348543.

[73] R. SPICA, P. ROBUFFO GIORDANO.*Active Decentralized Scale Estimation for Bearing-Based Localization*, in "IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, IROS'16", Daejeon, South Korea, October 2016, https://hal.inria.fr/hal-01355789.

[74] W. VAN TOLL, R. TRIESSCHEIJN, M. KALLMANN, R. OLIVA, N. PELECHANO, J. PETTRÉ, R. GERAERTS.*A comparative study of navigation meshes*, in "MIG '16 - 9th International Conference on Motion in Games", San Francisco, United States, ACM, October 2016, p. 91 - 100 [*DOI :* 10.1145/2994258.2994262], https://hal.inria.fr/hal-01392267.

[75] A. YOL, E. MARCHAND, F. CHAUMETTE, K. KANANI, T. CHABOT.*Vision-based navigation in low earth orbit*, in "Int. Symp. on Artificial Intelligence, Robotics and Automation in Space, i-SAIRAS'16", Beijing, China, June 2016, https://hal.inria.fr/hal-01304728.

Conferences without Proceedings

[76] S. BRIOT, F. CHAUMETTE, P. MARTINET.*Revisiting the determination of the singularity cases in the visual servoing of image points through the concept of "hidden robot"*, in "2017 IEEE International Conference on Robotics and Automation (ICRA 2017)", Singapour, Singapore, May 2017, https://hal.archives-ouvertes.fr/hal-01435810.

Scientific Books (or Scientific Book chapters)

[77] F. CHAUMETTE, S. HUTCHINSON, P. CORKE.*Visual Servoing*, in "Handbook of Robotics, 2nd edition", B. SICILIANO, O. KHATIB (editors), Springer, 2016, p. 841-866, https://hal.inria.fr/hal-01355384.

[78] M. OURAK, B. TAMADAZTE, N. ANDREFF, É. MARCHAND.*Multimodal Image Registration and Visual Servoing*, in "Lecture Notes in Electrical Engineering", Springer, July 2016, vol. 383, p. 157 - 175 [*DOI :* 10.1007/978-3-319-31898-1_9], https://hal.inria.fr/hal-01433996.

Research Reports

[79] S. BRIOT, P. MARTINET, F. CHAUMETTE.*Technical Report associated with the Paper: "Determining the Singularities for the Observation of Three Image Lines"*, IRCCyN ; IRISA, November 2016, https://hal.archives-ouvertes.fr/hal-01400575.

Other Publications

[80] D. J. AGRAVANTE, F. CHAUMETTE.*Combining visual servoing and walking in an acceleration resolved whole-body control framework*, June 2016, Journées Nationales de la Recherche Humanoïde, Toulouse, France, https://hal.inria.fr/hal-01358639.

[81] G. CLAUDIO, D. J. AGRAVANTE, F. SPINDLER, F. CHAUMETTE.*Dual arm manipulation and whole body control with the humanoid robot Romeo by visual servoing*, June 2016, Journées Nationales de la Recherche Humanoïde, Toulouse, France, https://hal.inria.fr/hal-01358124.

Project-Team LINKMEDIA

Creating and exploiting explicit links between multimedia fragments

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

IN PARTNERSHIP WITH:
CNRS

Institut national des sciences appliquées de Rennes

Université Rennes 1

RESEARCH CENTER
Rennes - Bretagne-Atlantique

THEME
Vision, perception and multimedia interpretation

Table of contents

Project-Team LINKMEDIA

Creation of the Project-Team: 2014 July 01

Keywords:

Computer Science and Digital Science:

3.3.2. - Data mining
3.3.3. - Big data analysis
5.3.3. - Pattern recognition
5.4.1. - Object recognition
5.4.3. - Content retrieval
5.7. - Audio modeling and processing
5.8. - Natural language processing
8.2. - Machine learning
8.4. - Natural language processing

Other Research Topics and Application Domains:

9. - Society and Knowledge

1. Members

Research Scientists

Guillaume Gravier [Team leader, CNRS, Senior Researcher, HDR]
Laurent Amsaleg [CNRS, Researcher, HDR]
Yannis Avrithis [Inria, Advanced Research position]
Vincent Claveau [CNRS, Researcher]
Teddy Furon [Inria, Researcher]

Faculty Members

Ewa Kijak [Univ. Rennes I, Associate Professor]
Simon Malinowski [Univ. Rennes I, Associate Professor]
Christian Raymond [INSA Rennes, Associate Professor]
Pascale Sébillot [INSA Rennes, Professor, HDR]

Technical Staff

Gabriel Sargent [CNRS]
Ronan Sicre [CNRS]
Arnaud Touboulic [CNRS, from Feb. 2016]

PhD Students

Raghavendran Balu [Inria, until Oct. 2016]
Rémi Bois [CNRS]
Ricardo Carlini Sperandio [PUC Minas]
Clément Dalloux [CNRS, from Dec. 2016]
Mikail Demirdelen [Univ. Rennes I, from Sep. 2016]
Ahmet Iscen [Inria]
Grégoire Jadi [Univ. Nantes]
Cédric Maigrot [Univ. Rennes I]
Oriane Simeoni [Inria, from Oct. 2016]
Vedran Vukotić [INSA Rennes]

Administrative Assistant
Aurélie Patier [Univ. Rennes I]

2. Overall Objectives

2.1. Context

Linked media appears today as a major challenge, with numerous potential applications in all areas of multimedia. The strong increase of ubiquitous access to the Internet and the resulting convergence of media on the network open countless opportunities for linked media and reinforce the key role of such a challenge. New applications centered on the notion of linked media are emerging today, such as second screen applications and recommendation services. However, because of the lack of adequate technology, linking related content is mostly deferred to human operators in current applications or to user behavior analysis, e.g., via collaborative filtering, thus indirectly considering the content. This fact severely limits the opportunities offered by a web of media, in terms of creativity, scalability, representativeness and completeness, thus negatively impacting the spread of linked media and the development of innovative services in the Internet of media.

Most of the research effort in automatic multimedia content analysis has been devoted so far to describing and indexing content on which core tasks around information retrieval and recommendation are built to develop multimedia applications. This general philosophy mostly reposes on a vision where documents are considered as isolated entities, i.e., as a basic unit which is indexed or analyzed regardless of other content items and of context. Considering documents in isolation has enabled key progress in content-based analysis and retrieval on a large scale: e.g., design of generic descriptors, efficient techniques for content-based analysis, fast retrieval methodology. But ignoring the links, implicit or explicit, between content items also appears as a rather strong assumption with direct consequences on algorithms and applications, both in terms of performance and in terms of possibilities.

2.2. Scientific objectives

LINKMEDIA investigates a number of key issues related to multimedia collections structured with explicit links: Can we discover what characterizes a collection and makes its coherence? Are there repeating motifs that create natural links and which deserve characterization and semantic interpretation? How to explicitly create links from pairwise distances? What structure should a linked collection have? How do we explain the semantic of a link? How explicit links can be used to improve information retrieval? To improve user experience? In this general framework, the global objective of LINKMEDIA is to develop the scientific, methodological and technological foundations facilitating or automating the creation, the description and the exploitation of multimedia collections structured with explicit links. In particular, we target a number of key contributions in the following areas:

- designing efficient methods dedicated to multimedia indexing and unsupervised motif discovery: efficiently comparing content items on a large scale and finding repeating motifs in an unsupervised manner are two key ingredients of multimedia linking based on a low-level representation of the content;
- improving techniques for structuring and semantic description: better description of multimedia content at a semantic—i.e., human interpretable—level, making explicit the implicit structure when it exists, is still required to make the most of multimedia data and to facilitate the creation of links to a precise target at a semantic level;
- designing and experimenting approaches to multimedia content linking and collection structuring: exploiting low-level and semantic content-based proximity to create explicit links within a collection requires specific methodology departing from pairwise comparison and must be confronted with real data;
- studying new paradigms for the exploitation of linked multimedia content as well as new usages: explicit links within media content collections change how such data is processed by machines and ultimately consumed by humans in ways that have yet to be invented and studied.

3. Research Program

3.1. Scientific background

LINKMEDIA is a multidisciplinary research team, with multimedia data as the main object of study. We are guided by the data and their specificity—semantically interpretable, heterogeneous and multimodal, available in large amounts, unstructured and disconnected—, as well as by the related problems and applications.

With multimedia data at the center, orienting our choices of methods and algorithms and serving as a basis for experimental validation, the team is directly contributing to the following scientific fields:

- multimedia: content-based analysis; multimodal processing and fusion; multimedia applications;
- computer vision: compact description of images; object and event detection;
- natural language processing: topic segmentation; information extraction;
- information retrieval: high-dimensional indexing; approximate k-nn search; efficient set comparison.

LINKMEDIA also takes advantage of advances in the following fields, adapting recent developments to the multimedia area:

- signal processing: image processing; compression;
- machine learning: deep architectures; structured learning; adversarial learning;
- security: data encryption; differential privacy;
- data mining: time series mining and alignment; pattern discovery; knowledge extraction.

3.2. Workplan

Research activities in LINKMEDIA are organized along three major lines of research which build upon the scientific domains already mentioned.

3.2.1. Unsupervised motif discovery

As an alternative to supervised learning techniques, unsupervised approaches have emerged recently in multimedia with the goal of discovering directly patterns and events of interest from the data, in a totally unsupervised manner. In the absence of prior knowledge on what we are interested in, meaningfulness can be judged based on one of three main criteria: unexpectedness, saliency and recurrence. This last case posits that repeating patterns, known as motifs, are potentially meaningful, leading to recent work on the unsupervised discovery of motifs in multimedia data [54], [52], [53].

LINKMEDIA seeks to *develop unsupervised motif discovery approaches which are both accurate and scalable*. In particular, we consider the discovery of repeating objects in image collections and the discovery of repeated sequences in video and audio streams. Research activities are organized along the following lines:

- developing the scientific basis for scalable motif discovery: sparse histogram representations; efficient co-occurrence counting; geometry and time aware indexing schemes;
- designing and evaluating accurate and scalable motif discovery algorithms applied to a variety of multimedia content: exploiting efficient geometry or time aware matching functions; fast approximate dynamic time warping; symbolic representations of multimedia data, in conjunction with existing symbolic data mining approaches;
- developing methodology for the interpretation, exploitation and evaluation of motif discovery algorithms in various use-cases: image classification; video stream monitoring; transcript-free natural language processing (NLP) for spoken document.

3.2.2. Description and structuring

Content-based analysis has received a lot of attention from the early days of multimedia, with an extensive use of supervised machine learning for all modalities [56], [48]. Progress in large scale entity and event recognition in multimedia content has made available general purpose approaches able to learn from very large data sets and performing fairly decently in a large number of cases. Current solutions are however limited to simple, homogeneous, information and can hardly handle structured information such as hierarchical descriptions, tree-structured or nested concepts.

LINKMEDIA aims at *expanding techniques for multimedia content modeling, event detection and structure analysis*. The main transverse research lines that LINKMEDIA will develop are as follows:

- context-aware content description targeting (homogeneous) collections of multimedia data: latent variable discovery; deep feature learning; motif discovery;
- secure description to enable privacy and security aware multimedia content processing: leveraging encryption and obfuscation; exploring adversarial machine learning in a multimedia context; privacy-oriented image processing;
- multilevel modeling with a focus on probabilistic modeling of structured multimodal data: multiple kernels; structured machine learning; conditional random fields.

3.2.3. Linking and collection data model

Creating explicit links between media content items has been considered on different occasions, with the goal of seeking and discovering information by browsing, as opposed to information retrieval via ranked lists of relevant documents. Content-based link creation has been initially addressed in the hypertext community for well-structured texts [47] and was recently extended to multimedia content [57], [51], [50]. The problem of organizing collections with links remains mainly unsolved for large heterogeneous collections of unstructured documents, with many issues deserving attention: linking at a fine semantic grain; selecting relevant links; characterizing links; evaluating links; etc.

LINKMEDIA targets pioneering research on media linking by *developing scientific ground, methodology and technology for content-based media linking* directed to applications exploiting rich linked content such as navigation or recommendation. Contributions are concentrated along the following lines:

- algorithmic of linked media for content-based link authoring in multimedia collections: time-aware graph construction; multimodal hypergraphs; large scale k-nn graphs;
- link interpretation and characterization to provide links semantics for interpretability: text alignment; entity linking; intention vs. extension;
- linked media usage and evaluation: information retrieval; summarization; data models for navigation; link prediction.

4. Application Domains

4.1. Asset management in the entertainement business

Regardless of the ingestion and storage issues, media asset management—archiving, describing and retrieving multimedia content—has turned into a key factor and a huge business for content and service providers. Most content providers, with television channels at the forefront, rely on multimedia asset management systems to annotate, describe, archive and search for content. So do archivists such as the Institut National de l'Audiovisuel, the Nederlands Instituut voor Beeld en Geluid or the British Broadcast Corporation, as well as media monitoring companies, such as Yacast in France. Protecting copyrighted content is another aspect of media asset management.

4.2. Multimedia Internet

One of the most visible application domains of linked multimedia content is that of multimedia portals on the Internet. Search engines now offer many features for image and video search. Video sharing sites also feature search engines as well as recommendation capabilities. All news sites provide multimedia content with links between related items. News sites also implement content aggregation, enriching proprietary content with user-generated content and reactions from social networks. Most public search engines and Internet service providers offer news aggregation portals.

4.3. Multiscreen TV

The convergence between television and the Internet has accelerated significantly over the past few years, with the democratization of TV on-demand and replay services and the emergence of social TV services and multiscreen applications. These evolutions and the consequently ever growing number of innovative applications offer a unique playground for multimedia technologies. Recommendation plays a major role in connected TV. Enriching multimedia content, with explicit links targeting either multimedia material or knowledge databases, appears as a key feature in this context, at the core of rich TV and second screen applications.

4.4. E-learning

On-line courses are rapidly gaining interest with the recent movement for massive open on-line courses (MOOCs). Such courses usually aggregate multimedia material, such as a video of the course with handouts and potentially text books, exercises and other related resources. This setting is very similar to that of the media aggregation sites though in a different domain. Automatically analyzing and describing video and textual content, synchronizing all material available across modalities, creating and characterizing links between related material or between different courses are all necessary features for on-line courses authoring.

5. Highlights of the Year

5.1. Highlights of the Year

LINKMEDIA ranked first at the TRECVid 2016 Hyperlinking international benchmark [12].

LINKMEDIA is selected as the organizer of the IEEE Workshop on Information Forensics and Security in 2017.

LINKMEDIA deeply involved in the winning bid for the organization of the ACM Conf. on Multimedia in 2019.

6. New Software and Platforms

6.1. A||GO multimedia web services

Participants: Guillaume Gravier [correspondent], Gabriel Sargent.

Available at http://allgo.irisa.fr, the A||GO platform allows for the easy deployment of the technology developed in the team as web services. Based on the A||GO infrastructure, LINKMEDIA has continued making available a number of web services related to multimedia content analysis. In 2016, we revamped the multimedia-related web service offer, making all serices interoperable and broadening the scope of services made available (silence detection, face detection, text-based fragmentation). We also made available short videos that demonstrate the usefulness of our web services.

7. New Results

7.1. Unsupervised motif and knowledge discovery

7.1.1. Multimodal person discovery in TV broadcasts

Participants: Guillaume Gravier, Gabriel Sargent, Ronan Sicre.

Work in collaboration with Silvio J. Guimarães, Gabriel B. de Fonseca and Izabela Lyon Freire, PUC Minas, in the framework of the Inria Associate Team MOTIF.

Pursuing efforts initiated in 2015 in the framework of the MediaEval benchmark on Multimodal Person Discovery, we investigated graph-based approaches to name the persons on screen and speaking in TV broadcasts with no prior information, leveraging text overlays, speech transcripts as well as face and voice comparison. We adopted a graph-based representation of speaking faces and investigated two tag-propagation approaches to associate overlays co-occurring with some speaking faces to other visually or audiovisually similar speaking faces. Given a video, we first build a graph from the detected speaking faces (nodes) and their audiovisual similarities (edges). Each node is associated to its co-occurring overlays (tags) when they exist. Then, we consider two tag-propagation approaches, respectively based on a random walk strategy and on Kruskal's minimum spanning tree algorithm for node clustering [28].

7.1.2. Efficient similarity self-join for near-duplicate video detection

Participants: Laurent Amsaleg, Guillaume Gravier.

Work in collaboration with Henrique B. da Silva, Silvio J. Guimarães, Zenilto do Patrocino Jr., PUC Minas, and Arnaldo de A. Araújo, UFMG, in the framework of the Inria Associate Team MOTIF.

The huge amount of redundant multimedia data, like video, has become a problem in terms of both space and copyright. Usually, the methods for identifying near-duplicate videos are neither adequate nor scalable to find pairs of similar videos. Similarity self-join operation could be an alternative to solve this problem in which all similar pairs of elements from a video dataset are retrieved. Methods for similarity self-join however exhibit poor performance when applied to high-dimensional data. In [33], we propose a new approximate method to compute similarity self-join in sub-quadratic time in order to solve the near-duplicate video detection problem. Our strategy is based on clustering techniques to find out groups of videos which are similar to each other.

7.1.3. Recommendation systems with matrix factorization

Participants: Raghavendran Balu, Teddy Furon.

Matrix factorization is a prominent technique for approximate matrix reconstruction and noise reduction. Its common appeal is attributed to its space efficiency and its ability to generalize with missing information. For these reasons, matrix factorization is central to collaborative filtering systems. In the real world, such systems must deal with million of users and items, and they are highly dynamic as new users and new items are constantly added. Factorization techniques, however, have difficulties to cope with such a demanding environment. Whereas they are well understood with static data, their ability to efficiently cope with new and dynamic data is limited. Scaling to extremely large numbers of users and items is also problematic. In [10], we propose to use the count sketching technique for representing the latent factors with extreme compactness, facilitating scaling.

In [11], we discovered that sketching techniques implicitly provide differential privacy guarantees thanks to the inherent randomness of the data structure. Collaborative filtering is a popular technique for recommendation system due to its domain independence and reliance on user behavior data alone. But the possibility of identification of users based on these personal data raise privacy concerns. Differential privacy aims to minimize these identification risks by adding controlled noise with known characteristics. The addition of noise impacts the utility of the system and does not add any other value to the system other than enhanced privacy.

7.2. Multimedia content description and structuring

7.2.1. Hierarchical topic structuring

Participants: Guillaume Gravier, Pascale Sébillot.

In [37], we investigated the potential of a topical structure of text-like data that we recently proposed [55] in the context of summarization and anchor detection in video hyperlinking. This structure is produced by an algorithm that exploits temporal distributions of words through word burst analysis to generate a hierarchy of topically focused fragments. The obtained hierarchy aims at filtering out non-critical content, retaining only the salient information at various levels of detail. For the tasks we choose to evaluate the structure on, the lost of important information is highly damaging. We show that the structure can actually improve the results of summarization or at least maintain state-of-the-art results, while for anchor detection it leads us to the best precision in the context of the Search and Anchoring in Video Archives task at MediaEval. The experiments were carried on written text and a more challenging corpus containing automatic transcripts of TV shows.

7.2.2. Multimedia-inspired descriptors for time series classification

Participant: Simon Malinowski.

The SIFT framework has shown to be effective in the image classification context. Recently, we designed a bag-of-words approach based on an adaptation of this framework to time series classification. It relies on two steps: SIFT-based features are first extracted and quantized into words; histograms of occurrences of each word are then fed into a classifier. In [38], we investigated techniques to improve the performance of bag-of-temporal-SIFT-words: dense extraction of keypoints and different normalizations of Bag-of-Words histograms. Extensive experiments have shown that our method significantly outperforms nearly all tested standalone baseline classifiers on publicly available UCR datasets. In [23], we also investigate the use of convolutional neural networks (CNN) for time series classification. Such networks have been widely used in many domains like computer vision and speech recognition, but only a little for time series classification. We have designed a convolutional neural network that consists of two convolutional layers. One drawback with CNN is that they need a lot of training data to be efficient. We propose two ways to circumvent this problem: designing data-augmentation techniques and learning the network in a semi-supervised way using training time series from different datasets. These techniques are experimentally evaluated on a benchmark of time series datasets.

7.2.3. Early time series classification

Participant: Simon Malinowski.

In time series classification, two antagonist notions are at stake. On the one hand, in most cases, the sooner the time series is classified , the higher the reward. On the other hand, an early classification is more likely to be erroneous. Most of the early classification methods have been designed to take a decision as soon as a sufficient level of reliability is reached. However, in many applications, delaying the decision with no guarantee that the reliability threshold will be met in the future can be costly. Recently, a framework dedicated to optimizing the trade-off between classification accuracy and the cost of delaying the decision was proposed, together with an algorithm that decides online the optimal time instant to classify an incoming time series. On top of this framework , we have built in [29] two different early classification algorithms that optimize the trade-off between decision accuracy and the cost of delaying the decision. These algorithms are non-myopic in the sense that, even when classification is delayed, they can provide an estimate of when the optimal classification time is likely to occur. Our experiments on real datasets demonstrate that the proposed approaches are more robust than existing methods.

7.3. Content-based information retrieval

7.3.1. Bi-directional embeddings for cross-modal content matching

Participants: Guillaume Gravier, Christian Raymond, Vedran Vukotić.

Common approaches to problems involving multiple modalities (classification, retrieval, hyperlinking, etc.) are early fusion of the initial modalities and crossmodal translation from one modality to the other. Recently, deep neural networks, especially deep autoencoders, have proven promising both for crossmodal translation and for early fusion via multimodal embedding. In [31], we propose a flexible cross-modal deep neural network architecture for multimodal and crossmodal representation. By tying the weights of two deep neural networks, symmetry is enforced in central hidden layers thus yielding a multimodal representation space common to the two original representation spaces. The proposed architecture is evaluated in multimodal query expansion and multimodal retrieval tasks within the context of video hyperlinking. In [32], we extend the approach, focusing on the evaluation of a good single-modal continuous representations both for textual and for visual information. word2vec and paragraph vectors are evaluated for representing collections of words, such as parts of automatic transcripts and multiple visual concepts, while different deep convolutional neural networks are evaluated for directly embedding visual information, avoiding the creation of visual concepts. We evaluate methods for multimodal fusion and crossmodal translation, with different single-modal pairs, in the task of video hyperlinking.

7.3.2. Intrinsic dimensions in language information retrieval

Participant: Vincent Claveau.

Examining the properties of representation spaces for documents or words in information retrieval (IR) brings precious insights to help the retrieval process. Recently, several authors have studied the real dimensionality of the datasets, called intrinsic dimensionality, in specific parts of these spaces. In [34], we propose to revisit this notion through a coefficient called α in the specific case of IR and to study its use in IR tasks. More precisely, we show how to estimate α from IR similarities and to use it in representation spaces used for documents and words. Indeed, we prove that α may be used to characterize difficult queries. We moreover show that this intrinsic dimensionality notion, applied to words, can help to choose terms to use for query expansion.

7.3.3. Evaluation of distributional thesauri

Participants: Vincent Claveau, Ewa Kijak.

With the success of word embedding methods, all the fields of distributional semantics have experienced a renewed interest. Beside the famous word2vec, recent studies have presented efficient techniques to build distributional thesaurus, including our work on information retrieval (IR) tools and concepts to build a thesaurus [14]. In [13], we address the problem of the evaluation of such thesauri or embedding models. Several evaluation scenarii are considered: direct evaluation through reference lexicons and specially crafted datasets, and indirect evaluation through a third party tasks, namely lexical subsitution and Information Retrieval. Through several experiments, we first show that the recent techniques for building distributional thesaurus outperform the word2vec approach, whatever the evaluation scenario. We also highlight the differences between the evaluation scenarii, which may lead to very different conclusions when comparing distributional models. Last, we study the effect of some parameters of the distributional models on these various evaluation scenarii.

7.3.4. Scaling group testing similarity search

Participants: Laurent Amsaleg, Ahmet Iscen, Teddy Furon.

The large dimensionality of modern image feature vectors, up to thousands of dimensions, is challenging high dimensional indexing techniques. Traditional approaches fail at returning good quality results within a response time that is usable in practice. However, similarity search techniques inspired by the group testing framework have recently been proposed in an attempt to specifically defeat the curse of dimensionality. Yet, group testing does not scale and fails at indexing very large collections of images because its internal procedures analyze an excessively large fraction of the indexed data collection. In [16], we identify these difficulties and proposes extensions to the group testing framework for similarity searches that allow to handle larger collections of feature vectors. We demonstrate that it can return high quality results much faster compared to state-of-the-art group testing strategies when indexing truly high-dimensional features that are indeed hardly indexable with traditional indexing approaches.

We also discovered that group testing helps in enforcing security and privacy in identification. We detail a particular scheme based on embedding and group testing. Whereas embedding poorly protects the data when used alone, the group testing approach makes it much harder to reconstruct the data when combined with embedding. Even when curious server and user collude to disclose the secret parameters, they cannot accurately recover the data. Our approach reduces as well the complexity of the search and the required storage space. We show the interest of our work in a benchmark biometrics dataset [17], where we verify our theoretical analysis with real data.

7.3.5. Large-scale similarity search using matrix factorization

Participants: Ahmet Iscen, Teddy Furon.

Work in collaboration with Michael Rabbat, McGill University, Montréal.

We consider the image retrieval problem of finding the images in a dataset that are most similar to a query image. Our goal is to reduce the number of vector operations and memory for performing a search without sacrificing accuracy of the returned images. In [18], we adopt a group testing formulation and design the decoding architecture using either dictionary learning or eigendecomposition. The latter is a plausible option for small-to-medium sized problems with high-dimensional global image descriptors, whereas dictionary learning is applicable in large-scale scenarios. Experiments with standard image search benchmarks, including the Yahoo100M dataset comprising 100 million images, show that our method gives comparable (and sometimes better) accuracy compared to exhaustive search while requiring only 10 % of the vector operations and memory. Moreover, for the same search complexity, our method gives significantly better accuracy compared to approaches based on dimensionality reduction or locality sensitive hashing.

7.4. Linking, navigation and analytics

7.4.1. Opinion similarity and target extraction

Participants: Vincent Claveau, Grégoire Jadi.

Work in collaboration with Laura Monceaux and Béatrice Daille, LINA, Nantes.

In [19], we propose to evaluate the lexical similarity information provided by word representations against several opinion resources using traditional information retrieval tools. Word representation have been used to build and to extend opinion resources, such as lexicon and ontology, and their performance have been evaluated on sentiment analysis tasks. We question this method by measuring the correlation between the sentiment proximity provided by opinion resources and the semantic similarity provided by word representations using different correlation coefficients. We also compare the neighbors found in word representations and list of similar opinion words. Our results show that the proximity of words in state-of-the-art word representations is not very effective to build sentiment similarity.

In [20], we present the development of an opinion target extraction system in English and transpose it to the French language. In addition, we realize an analysis of the features and their effectiveness in English and French which suggest that it is possible to build an opinion target extraction system independant of the domain. Finally, we propose a comparative study of the errors of our systems in both English and French and propose several solutions to these problems.

7.4.2. Reinformation and fake detection in social networks

Participants: Vincent Claveau, Ewa Kijak, Cédric Maigrot.

Traditional media are increasingly present on social networks, but these usual sources of information are confronted with other sources called reinformation sources. These last ones sometimes tend to distort the information relayed to match their ideologies, rendering it partially or totally false. In [25], we conduct a study pursuing two goals: first, we present a corpus containing Facebook messages issued from both types of media sources; secondly, we propose some experiments in order to automatically detect reinformation messages. In particular, we investigate the influence of shallow features versus features more specifically describing the message content. We also developed a multi-modal hoax detection system composed of text, source, and image analysis [24]. As hoax can be very diverse, we want to analyze several modalities to better detect them. This system is applied in the context of the Verifying Multimedia Use task of MediaEval 2016. Experiments show the performance of each separated modality as well as their combination.

7.4.3. Multimodal video hyperlinking

Participants: Rémi Bois, Guillaume Gravier, Christian Raymond, Pascale Sébillot, Ronan Sicre, Vedran Vukotić.

Pursuing previous work on video hyperlinking and recent advances in multimodal content matching [32], we benchmarked a full video hyperlinking system in the framework of the TRECVid international benchmark [12]. The video hyperlinking task aims at proposing a set of video segments, called targets, to complement a query video segment defined as anchor. The 2016 edition of the task encouraged participants to use multiple modalities. In this context, we chose to submit four runs in order to assess the pros and cons of using two modalities instead of a single one and how crossmodality differs from multimodality in terms of relevance. The crossmodal run performed best and obtained the best precision at rank 5 among participants. In parallel, we also demonstrated that, in this framework, multimodal and crossmodal approaches offer significantly more diversity in the set of target proposed than classical information retrieval based approaches where all modalities are combined. We compared bidirectionnal multimodal embeddings [31] with multimodal LDA approaches as experimented last year in TRECVid [49]. The former offers more accurate matching, the latter exhibiting slighlty more diveristy.

7.4.4. User-centric evaluation of hyperlinked news content

Participants: Rémi Bois, Guillaume Gravier, Pascale Sébillot, Arnaud Touboulic.

Work in collaboration with Éric Jamet, Martin Ragot and Maxime Robert, CRPCC, Rennes.

Following our study of professional user needs in multimedia news analytics [15], we developed a prototype news analytics interface that facilitates the exploration of collections of multimedia documents by journalists. The application, based on standard web technology, enriches classical functionalities for this type of applications (e.g., keyword highlights, named entity detection, keyword search, etc.) with navigation-based functionalities. The latter exploit a graph-based organization of the collection, established from content-based similarity graphs on which community detection is performed along with basic link characterization. We performed usage tests on students in journalism and on journalists where each user was asked to write a synthesis article on a given topic. Preliminary results indicate that the graph-based navigation improves the completeness of the synthesis by exposing users to more content than with a standard search engine.

7.5. Miscellaneous

In parallel with mainstream research activities, LINKMEDIA has a number of contribtions in other domains based on the expertise of the team members.

7.5.1. Bidirectional GRUs in spoken dialog

Participants: Christian Raymond, Vedran Vukotić.

Recurrent neural networks recently became a very popular choice for spoken language understanding (SLU) problems. They however represent a big family of different architectures that can furthermore be combined to form more complex neural networks. In [30], we compare different recurrent networks, such as simple recurrent neural networks, long short-term memory networks, gated memory units and their bidirectional versions, on the popular ATIS dataset and on MEDIA, a more complex French dataset. Additionally, we propose a novel method where information about the presence of relevant word classes in the dialog history is combined with a bidirectional gated recurrent unit (GRU).

7.5.2. Kernel principal components analysis with extreme learning machines

Participant: Christian Raymond.

Work in collaboration with M'Sila University, Algeria.

Nowadays, wind power and precise forecasting are of great importance for the development of modern electrical grids. In [26], we investigate a prediction system for time series based on kernel principal component analysis (KPCA) and extreme learning machine (ELM). Comparison with standard dimensionality reduction techniques show that the reduction of the original input space affects positively the prediction output.

7.5.3. Pronunciation adaptation for spontaneous speech synthesis

Work in collaboration with Gwénolé Lecorvé and Damien Lolive, IRISA, Rennes.

In [36], we present a new pronunciation adaptation method which adapts canonical pronunciations to a spontaneous style. This is a key task in text-to-speech as those pronunciation variants bring expressiveness to synthetic speech, thus enabling new potential applications. The strength of the method is to solely rely on linguistic features and to consider a probabilistic machine learning framework, namely conditional random fields, to produce the adapted pronunciations.

7.6. Participation in benchmarking initiatives

- Video hyperlinking, TRECVid
- Search and anchoring, Mediaeval Multimedia International Benchmark
- Multimodal person discovery in broadcast TV, Mediaeval Multimedia International Benchmark
- DeFT 2015 text-mining challenge

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Contracts with Industry

Teddy Furon spent 20 % of his time during 6 months to transfer research result to IRT B-com

Ph. D. contract with Alcatel-Lucent Bell Labs (Raghavendran Balu) in the framework of the joint Inria-Alcatel Lucent lab.

9. Partnerships and Cooperations

9.1. Regional Initiatives

9.1.1. CominLabs Project Linking Media in Acceptable Hypergraphs (LIMAH)

Participants: Rémi Bois, Vincent Claveau, Guillaume Gravier, Grégoire Jadi, Pascale Sébillot, Arnaud Touboulic.

Duration: 4 years, started in April 2014
Partners: Telecom Bretagne (IODE), Univ. Rennes II (CRPCC, PREFics), Univ. Nantes (LINA/TAL)
URL: http://limah.irisa.fr

LIMAH aims at exploring hypergraph structures for multimedia collections, instantiating actual links reflecting particular content-based proximity—similar content, thematic proximity, opinion expressed, answer to a question, etc. Exploiting and developing further techniques targeting pairwise comparison of multimedia contents from an NLP perspective, LIMAH addresses two key issues: How to automatically build from a collection of documents an hypergraph, i.e., a graph combining edges of different natures, which provides exploitable links in selected use cases? How collections with explicit links modify usage of multimedia data in all aspects, from a technology point of view as well as from a user point of view? LIMAH studies hypergraph authoring and acceptability taking a multidisciplinary approach mixing ICT, law, information and communication science as well as cognitive and ergonomy psychology.

9.1.2. CominLabs Project BigCLIN

Participants: Vincent Claveau, Ewa Kijak, Clément Dalloux.

Duration: 3 years, started in September 2016
Partners: STL-CNRS, Inserm/CHU Rennes, Inria Cidre URL: http://www.bigclin.cominlabs.ueb.eu

Data collected or produced during clinical care process can be exploited at different levels and across different domains. Yet, a well-known challenge for secondary use of health big data is that much of detailed patient information is embedded in narrative text, mostly stored as unstructured data. The project proposes to address the essential needs when reusing unstructured clinical data at a large scale. We propose to develop new clinical records representation relying on fine-grained semantic annotation thanks to new NLP tools dedicated to French clinical narratives. To efficiently map this added semantic information to existing structured data for further analysis at big scale, the project also addresses distributed systems issues: scalability, management of uncertain data and privacy, stream processing at runtime, etc.

9.2. National Initiatives

9.2.1. ANR Project IDFRAud

Participant: Teddy Furon.

Duration: 3 years, started in Feb. 2015
Partners: AriadNext, IRCGN, École Nationale Supérieure de Police

The IDFRAud project consists in proposing an automatic solution for ID analysis and integrity verification. Our ID analysis goes through three processes: classification, text extraction and ID verification. The three processes rely on a set of rules that are externalized in formal manner in order to allow easy management and evolving capabilities. This leads us to the ID knowledge management module. Finally, IDFRAud addresses the forensic link detection problem and to propose an automatic analysis engine that can be continuously applied on the detected fraud ID database. Cluster analysis methods are used to discover relations between false IDs in their multidimensional feature space. This pattern extraction module will be coupled with a suitable visualization mechanism in order to facilitate the comprehension and the analysis of extracted groups of inter-linked fraud cases.

9.2.2. FUI 19 NexGenTV

Participants: Vincent Claveau, Guillaume Gravier, Ewa Kijak, Gabriel Sargent, Ronan Sicre.

Duration: 2.5 years, started in May 2015
Partners: Eurecom, Avisto Telecom, Wildmoka, Envivio-Ericsson

Television is undergoing a revolution, moving from the TV screen to multiple screens. Today's user watches TV and, at the same time, browses the web on a tablet, sends SMS, posts comments on social networks, searches for complementary information on the program, etc. Facing this situation, NexGen-TV aims at developing a generic solution for the enrichment, the linking and the retrieval of video content targeting the cost-cutting edition of second screen and multiscreen applications for broadcast TV. The main outcome of the project will be a software platform to aggregate and distribute video content via a second-screen edition interface connected to social media. The curation interface will primarily make use of multimedia and social media content segmentation, description, linking and retrieval. Multiscreen applications will be developed on various domaine, e.g., sports, news.

9.3. European Initiatives

9.3.1. Collaborations with Major European Organizations

Big Data Value Association (BDVA)

LINKMEDIA is a co-founder and co-leader of the media group (TF7) within BDVA

9.4. International Initiatives

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

9.4.1.1. MOTIF

Title: Unsupervised motif discovery in multimedia content

International Partner (Institution - Laboratory - Researcher):

Pontifícia Universidade Católica de Minas Gerais (Brazil) - Audio-Visual Information Processing Laboratory (VIPLAB) - Silvio Jamil Guimarães

Universidade Federal Minas Gerais, Brasil - NPDI - Arnaldo Albuquerque de Araújo

Duration: 2014–2016

MOTIF aims at studying various approaches to unsupervised motif discovery in multimedia sequences, i.e., to the discovery of repeated sequences with no prior knowledge on the sequences. On the one hand, we will develop symbolic approaches inspired from work on bioinformatics to motif discovery in the multimedia context, investigating symbolic representations of multimedia data and adaptation of existing symbolic motif discovery algorithms. On the other hand, we will further develop cross modal clustering approaches to repeated sequence discovery in video data, building upon previous work.

9.4.2. Inria International Partners

9.4.2.1. Informal International Partners

- National Institute for Informatics, Japan
- University of Amsterdam, The Netherlands
- Czech Technical University, Czech Republic
- Katholieke Universiteit Leuven, Belgium

9.4.3. Participation in Other International Programs

- PICS CNRS MM-Analytics
  - Title: Fouille, visualisation et exploration multidimensionnelle de contenus multimédia ; Multi-Dimensional Multimedia Browsing, Mining, Analytics (num 6382).
  - International Partner (Institution - Laboratory - Researcher):

    Reykjavík University, Iceland - Björn Þór Jónsson
  - Jan. 2014 – Dec. 2016
- CNRS – CONFAP FIGTEM
  - Title: Fine-grained text-mining for clinical trials
  - International Partner (Institution - Laboratory - Researcher): Pontifícia Universidade Católica do Paraná - Health Informatics dept, Claudia Moro

    FIGTEM aims at developing natural language processing methods, including information extraction and indexing, dedicated to the clinical trial domain. The goal is to populate a formal representation of patients (via their electronic patient records) and clinical trial data in different languages (French, English, Portuguese).

– Jan. 2016 – Dec. 2018

9.5. International Research Visitors

9.5.1. Visits of International Scientists

Claudia Moro, Lucas Oliveira

Date: Oct. 2016 (1 week)

Institution: Pontifícia Universidade Católica do Paraná - Health Informatics dept

Giorgos Tolias

Date: Sept. 2016 (1 week)

Institution: Czech Technical University, Czech Republic

9.5.1.1. Internships

Gabriel B. de Fonseca

Date: Nov. 2016 - Jan. 2017

Institution: PUC Minas, Brazil

9.5.2. Visits to International Teams

Vincent Claveau

Date: 7-17 December 2016

Institution: Helth Informatics dept, Pontifícia Universidade Católica do Paraná, Curitiba, Brazil

Vincent Claveau

Date: 7-13 May 2016

Institution: OLST, Univ. of Montreal, Canada

Guillaume Gravier, Simon Malinowski

Date: Jul. 2016 (1 week)

Institution: PUC Minas, Brazil

Ahmet Iscen

Date: Apr. 2016 - May 2016

Institution: Czech Technical University, Czech Republic

Vedran Vukotić

Date: Sep. 2016 - Dec. 2016

Institution: TU Delft, The Netherlands

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events Organisation

10.1.1.1. General Chair, Scientific Chair

Laurent Amsaleg was the technical program chair of the Intl. Conf. on Similarity Search and Applications 2016.

Laurent Amsaleg and Vincent Claveau, with X. Tannier (LIMSI-CNRS), D. Nouvel (Inalco) organized a workshop about Computational Journalism. More than 50 persons (computer scientists, sociologists, and news/media compan fellows) attented. Program available at http://compjournalism2016.irisa.fr.

Pascale Sébillot is a member of the permanent steering committee of Conf. Francophone en Traitement Automatique des Langues Naturelles.

10.1.1.2. Member of the Organizing Committees

Simon Malinowski was in the organizing committee of the Workshop on Advanced Analysis and Learning on Temporal Data (AALTD 16) held in Riva del Garda, colocated with ECML/PKDD 2016.

Guillaume Gravier was co-chair of a special session at The Speaker and Language Recognition Workshop Odyssey 2016.

10.1.2. Scientific Events Selection

10.1.2.1. Chair of Conference Program Committees

Vincent Claveau was area chair of the French-speaking conference Traitement Automatique du Langage Naturel 2016.

Guillaume Gravier was technical program chair of the International Workshop on Content-Based Multimedia Indexing 2016.

Guillaume Gravier was area chair of the ACM Intl. Conf. on Multimedia 2016.

Guillaume Gravier was technical program chair of the French-speaking international conference Journées d'Études sur la Parole 2016.

Pascale Sébillot was area chair of the French-speaking conference Traitement Automatique du Langage Naturel 2016.

10.1.2.2. Member of Conference Program Committees

Laurent Amsaleg was a PC member of: Intl. Conf. on Similarity Search and Applications; ACM Intl. Conf. on Multimedia Retrieval; IEEE Intl. Conf. on Multimedia and Exhibition.

Yannis Avrithis was a PC member of: IEEE Conference on Computer Vision and Pattern Recognition; European Conference on Computer Vision; British Machine Vision Conference; ACM Multimedia Conference on Multimedia; European Signal Processing Conference; Intl. Conference on Multimedia and Expo; IEEE Intl. Conference in Image Processing; IEEE Intl. Conference on Acoustics, Speech, and Signal Processing; Intl. Workshop on Content-Based Multimedia Indexing.

Vincent Claveau was a PC member of: ACL demonstration track; Workshop on Content-Based Multimedia Indexing; Conf. Traitement Automatique du Langage Naturel; CORIA; Language Resources and Evaluation Conf.

Teddy Furon was a PC member of ACM Information Hiding and Multimedia Security.

Guillaume Gravier was a PC member of: ACM Intl. Conf. on Multimedia; IEEE Intl. Conf. on Multimedia and Exhibition; Annual Conf. of the Intl. Speech Communication Association; IEEE Intl. Workshop on Multimedia Signal Processing; European Conf. on Information Retrieval; Conf. on Knowledge and Information Management; IEEE Intl. Conf. on Acoustics, Speech and Signal Processing; Intl. Conf. on Statistical Language and Speech Processing; Language Resources and Evaluation Conf.

Ewa Kijak was a PC member of Intl. Workshop on Content-Based Multimedia Indexing.

Pascale Sébillot was a PC member of: ACM International Conference on Information and Knowledge Management; European Conference on Information Retrieval; Conférence nationale en intelligence artificielle; Conf. Traitement Automatique des Langues Naturelles; Colloque international sur l'écrit et le document

10.1.2.3. Reviewer

Ewa Kijak reviewed for ACM Conf. on Multimedia.

Pascale Sébillot was a scientific committee member of: Language Resources and Evaluation Conference

10.1.3. Journal

10.1.3.1. Member of the Editorial Boards

Vincent Claveau is member of the editoral board of the journal Traitement Automatique des Langues.

Vincent Claveau was guest editor of the special issue on Information Retrieval and Natural Language Processing of the journal Traitement Automatique des Langues (published in 2016).

Guillaume Gravier was appointed associate editor of IEEE Trans. on Multimedia in Sep. 2016.

Guillaume Gravier is guest editor of the special issue of Multimedia Tools and Applications on content-based multimedia indexing (to appear in 2017).

Guillaume Gravier was editor of the Working Notes Proc. of the MediaEval Multimedia Benchmark.

Pascale Sébillot is editor of the Journal Traitement Automatique des Langues.

Pascale Sébillot is member of the editorial board of the Journal Traitement Automatique des Langues.

10.1.3.2. Reviewer - Reviewing Activities

Laurent Amsaleg was a reviewer for: IEEE Trans. on Circuits and Systems for Video Technology, IEEE Trans. on Multimedia, Pattern Recognition Letters, IEEE Trans. on Information Forensics & Security, IEEE Trans. on Image Processing.

Yannis Avrithis was a reviewer for Multimedia Tools and Applications, Elsevier Neurocomputing.

Vincent Claveau reviewed for Multimedia Tools and Applications, Elsevier Knowledge-Based Systems, Elsevier Information Sciences, Traitement Automatique des Langues.

Teddy Furon is a reviewer for IEEE Trans. on Information Forensics and Security, Elsevier Digital Signal Processing, EURASIP Journal of Information Security, IEEE Trans. on Image Processing, IEEE Trans. on Signal Processing.

Guillaume Gravier reviewed for Computer Speech and Language.

Pascale Sébillot was member of the reading committee for several issues of the Journal Traitement Automatique des Langues.

10.1.4. Invited Talks

Yannis Avrithis gave an invited talk at Pattern Recognition and Computer Vision Colloquium on March 31 in Prague, http://cmp.felk.cvut.cz/cmp/events/colloquium-2016.03.31.

Vincent Claveau gave an invited talk at Congrès de l'ACFAS, colloque Documents et ressources pour leur traitement : un couplage crucial, Montreal.

Vincent Claveau gave an invited talk at Séminaire D2K du labex Digicosme, Paris.

Vincent Claveau gave an invited talk at Séminaire commun RALI-OLST, Montréal.

Guillaume Gravier gave an invited talk at the ACM Multimedia 2016 TPC workshop.

Guillaume Gravier was invited as a panelist at the Intl. Workshop on Content-Based Multimedia Indexing.

Guillaume Gravier gave an invited talk at the Journées de Statistique de Rennes (JSTAR).

Guillaume Gravier animated a round-table on 'Jeux de données et données en jeu : l'accès des données à des fins de recherche' at the Rencontres du Numérique organized by the National Research Agency (ANR).

Simon Malinowski gave an invited talk on time series classification at PUC Minas, Brazil.

Pascale Sébillot gave an invited tutorial (8h) about NLP at Institut de l'information scientifique et technique (INiST) Nancy.

10.1.5. Leadership within the Scientific Community

Laurent Amsaleg was nominated to be part of the Steering Committee of SISAP for a 2016–2020 term.

Vincent Claveau is finance head of the Association pour la Recherche d'Informations et ses Applications (ARIA).

Guillaume Gravier is president of the Association Francophone de la Communication Parlée (AFCP), French-speaking branch of the Intl. Speech Communication Association.

Guillaume Gravier is co-founder and general chair of the ISCA SIG Speech, Language and Audio in Multimedia.

Guillaume Gravier is member of the Community Council of the Mediaeval Multimedia Evaluation series.

Guillaume Gravier is the technical representative of Inria in the cPPP Big Data Value Association, actively working on technical aspects of data analytics.

10.1.6. Scientific Expertise

Laurent Amsaleg was an Expert Evaluator for the Luxembourg National Research Fund, 2016.

Vincent Claveau was an expert for the FNRS (Belgian funding agency).

Vincent Claveau was an expert for GIFAS - Groupement des Industries Françaises Aéronautiques et Spatiales.

Vincent Claveau was an expert for the Programme Hubert Curien (funding program for bilateral projects).

Teddy Furon is scientific adviser for the company LAMARK.

Guillaume Gravier was vice-president of the Scientific Evaluation Committee of the National Research Agency for the theme 'Interaction, Robotic, Content, Automatic' up to July 2016.

Since September 2016, Guillaume Gravier is president of the Scientific Evaluation Committee of the National Research Agency for the theme 'Knowledge, data, content, big data, HPC, simulation'.

10.1.7. Research Administration

Vincent Claveau is deputy head of the GdR MaDICS, a CNRS inter-lab initiative to promote research about Big Data and Data Science.

Guillaume Gravier is a member of the Board of the technology cluster Images & Réseaux.

Guillaume Gravier is a member of the Board of the Comité des Projets of Inria - Rennes Bretagne Atlantique.

Pascale Sébillot is a member of the Conseil National des Universités 27th section (computer science).

Pascale Sébillot is the director of the Computer Science Laboratory, INSA Rennes.

Pascale Sébillot is the deputy director of the Scientific Advisory Committee of IRISA UMR 6074.

Pascale Sébillot is a member of the theses advisory committee of the Matisse doctoral school.

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

For reseachers, all activities are given. For professors and assistant professors, only courses at the M. Sc. level are listed.

Licence: Teddy Furon, Probabilities, 40h, L1, Agrocampus Rennes, France

Licence: Guillaume Gravier, Databases, 30h, L2, INSA Rennes, France

Licence: Guillaume Gravier, Probability and statistics, 16h, L3, INSA Rennes, France

Licence: Guillaume Gravier, Natural Language Processing, 12h, L3, INSA Rennes, France

Master: Laurent Amsaleg, Multidimensional indexing, 13h, M2, University Rennes 1, France

Master: Vincent Claveau, Information retrieval, 7h, M2, ENSAI, Rennes, France

Master: Vincent Claveau, Data-Based Knowledge Acquisition: Symbolic Methods, 20h, M1, INSA de Rennes, France

Master: Vincent Claveau, Text Mining, 36h, M2, Univ. Rennes 1, France

Master: Vincent Claveau, Machine Learning for symbolic and sequential data, 7h, M2, Univ. Rennes 1, France

Master: Vincent Claveau, Information Retrieval, 15h, M2, ENSSAT, France

Master: Vincent Claveau, Information Retrieval, 13h, M2, Univ. Rennes 1, France

Master: Teddy Furon, Rare events, 20h, M2, Insa Rennes, France

Master: Guillaume Gravier, Data analysis and probabilistic modeling, 30h, M2, University Rennes 1, France

Master: Ewa Kijak, Image processing, 64h, M1, ESIR, France

Master: Ewa Kijak, Supervised learning, 15h, M2R, University Rennes 1, France

Master: Ewa Kijak, Supervised learning, 45h, M1, ESIR, France

Master: Ewa Kijak, Statistical data mining, 14h, M2, University Rennes 1, France

Master: Ewa Kijak, Indexing and multimedia databases, 15h, M2, ENSSAT, France

Master: Ewa Kijak, Computer vision, 15h, M2, ESIR, France

Master: Simon Malinowski, Short-term time series prediction, 29h, M1, Univ. Rennes 1

Master: Simon Malinowski, Supervised Learning, 10h, M2, Univ. Rennes 1

Master: Pascale Sébillot, Data-Based Knowledge Acquisition: Symbolic Methods, 14h, M1, INSA Rennes, France

Master: Pascale Sébillot, From multimedia documents structuring to multimedia collections structuring, 2h, M2 Language Industry, Université Grenoble-Alpes, France

Master: Pascale Sébillot, Advanced Databases and Modern Information Systems, 70h, M2, INSA Rennes, France

Master: Pascale Sébillot, Logic Programming, 12h, M1, INSA Rennes, France

10.2.2. Supervision

PhD: Raghavendran Balu, Privacy-preserving data aggregation and service personalization using highly-scalable data indexing techniques, defended Nov. 2016, Teddy Furon and Laurent Amsaleg

PhD: Petra Bosilj, Content based image indexing and retrieval using hierarchical image representations, defended Jan. 2016, Ewa Kijak and Sebastien Lefèvre (with OBELIX, IRISA team)

PhD: Thomas Gaillat, Reference in interlanguage: The case of this and that, defended Jun. 2016, Nicolas Ballier and Pascale Sébillot (with Université Paris Diderot)

PhD in progress: Rémi Bois, Navigable directed multimedia hypergraphs: construction and exploitation, started October 2014, Guillaume Gravier and Pascale Sébillot

PhD in progress: Ricardo Carlini Sperandio, Unsupervised motif mining in multimedia time series, started August 2015, Laurent Amsaleg and Guillaume Gravier

PhD in progress: Ahmet Iscen, Continuous memories for representing sets of vectors and image collections, started September 2014, Teddy Furon

PhD in progress: Grégoire Jadi, Opinion mining in multimedia data, started October 2014, Vincent Claveau, Béatrice Daille (LINA, Nantes) and Laura Monceaux (LINA, Nantes)

PhD in progress: Raheel Kareem Qader, Phonology modeling for emotional speech synthesis, started January 2014, Gwénolé Lecorvé and Pascale Sébillot (with EXPRESSION, IRISA Team)

PhD in progress: Mathieu Laroze, Active learning on adaptive representations for object detection in high-resolution imaging, started June 2016, Romain Dambreville, Chloe Friguet, Ewa Kijak and Sebastien Lefevre (with OBELIX, IRISA team)

PhD in progress: Cédric Maigrot, Detecting fake information on social networks, started October 2015, Laurent Amsaleg, Vincent Claveau and Ewa Kijak

PhD in progress: Vedran Vukotič, Deep neural architectures for automatic representation learning from multimedia multimodal data, started October 2014, Guillaume Gravier and Christian Raymond

PhD in progress: Oriane Simeoni, Invariance and supervision in visual learning, started Oct. 2016, Yannis Avrithis and Guillaume Gravier

PhD in progress: Mikail Demirdelen, User-adapted multi-document multimedia synthesis, started Oct. 2016, Guillaume Gravier and Pascale Sébillot

PhD in progress: Clément Dalloux, Clinical Text Mining and Indexing, started Dec. 2016, Olivier Dameron and Vincent Claveau (with DYLISS projet-team)

10.2.3. Juries

Vincent Claveau

PhD, reviewer, Gabriel Bernier-Colborne, Université de Montréal, Canada

Teddy Furon

PhD, reviewer, Thijs Laarhoven, Eindhoven University

PhD, jury member, Julien Lolive, Télécom-Bretagne Brest

Guillaume Gravier

PhD, reviewer and president, Xavier Bost, Université d'Avignon et des Pays du Vaucluse

Ewa Kijak

PhD, Hassan Wehbe, Université Toulouse III

Pascale Sébillot

PhD, reviewer, Jean-Philippe Fauconnier Biard, Université Toulouse 3

PhD, member, Abdessalam Bouchekif, Université du Maine

PhD, member, Sana Ben Nasr, Université de Rennes 1

HDR, member, Olivier Ferret, Université Paris Sud 11

10.3. Popularization

Vincent Claveau

An article entitled 'Vélos, bicyclettes et moteur de recherche', was published by ACFAS on its website

An article entitled 'Détecter l'intox sur Twitter' was published in a special issue of the Dossiers de l'Université de Rennes 1

11. Bibliography

Publications of the year

Doctoral Dissertations and Habilitation Theses

[1] R. BALU.*Privacy-aware and scalable recommender systems using sketching techniques*, Université Rennes 1, November 2016, https://tel.archives-ouvertes.fr/tel-01430156.

Articles in International Peer-Reviewed Journal

[2] P. BOSILJ, E. APTOULA, S. LEFÈVRE, E. KIJAK.*Retrieval of Remote Sensing Images with Pattern Spectra Descriptors*, in "ISPRS International Journal of Geo-Information", 2016 [*DOI :* 10.3390/IJGI5120228], https://hal.archives-ouvertes.fr/hal-01397883.

[3] P. BOSILJ, M. H. WILKINSON, E. KIJAK, S. LEFÈVRE.*Local 2D pattern spectra as connected region descriptors*, in "Mathematical Morphology - Theory and Applications", 2016 [*DOI :* 10.1515/MATHM-2016-0011], https://hal.archives-ouvertes.fr/hal-01320009.

[4] M. JAIN, H. JÉGOU, P. BOUTHEMY.*Improved Motion Description for Action Classification* , in "Frontiers in information and communication technologies", January 2016 [*DOI :* 10.3389/FICT.2015.00028], https://hal.inria.fr/hal-01401833.

[5] G. SARGENT, F. BIMBOT, E. VINCENT.*Estimating the structural segmentation of popular music pieces under regularity constraints*, in "IEEE/ACM Transactions on Audio, Speech, and Language Processing", 2017, https://hal.inria.fr/hal-01403210.

[6] L. WENG, L. AMSALEG, T. FURON.*Privacy-Preserving Outsourced Media Search*, in "IEEE Transactions on Knowledge and Data Engineering", July 2016, vol. 28, n^{o} 10 [*DOI :* 10.1109/TKDE.2016.2587258], https://hal.inria.fr/hal-01391444.

[7] C. E. DOS SANTOS, E. E. KIJAK, G. GRAVIER, W. ROBSON SCHWARTZ.*Partial Least Squares for Face Hashing*, in "Neurocomputing", 2016, vol. 213, p. 34-47 [*DOI :* 10.1016/J.NEUCOM.2016.02.083], https://hal.archives-ouvertes.fr/hal-01399660.

Articles in National Peer-Reviewed Journal

[8] V. CLAVEAU, J.-Y. NIE.*Recherche d'information et traitement automatique des langues : collaboration, synergie et convergence : Introduction au numéro spécial sur la recherche d'information et le traitement automatique des langues*, in "Traitement Automatique des Langues", August 2016, vol. 56, n^{o} 3, https://hal.archives-ouvertes.fr/hal-01394789.

Articles in Non Peer-Reviewed Journal

[9] I. BOGDAN, H. MÜLLER, G. GRAVIER, Y. KOMPATSIARIS.*2016 14th International Workshop on Content-Based Multimedia Indexing: IEEE Multimedia Conference Report*, in "IEEE MultiMedia", 2016, https://hal.archives-ouvertes.fr/hal-01399468.

International Conferences with Proceedings

[10] R. BALU, T. FURON, L. AMSALEG.*Sketching techniques for very large matrix factorization*, in "ECIR 2016 - 38th European Conference on Information Retrieval", Padoue, Italy, Proceedings of the European conference on Information Retrieval, 2016, https://hal.inria.fr/hal-01249621.

[11] R. BALU, T. FURON.*Differentially Private Matrix Factorization using Sketching Techniques*, in "ACM Workshop on Information Hiding and Multimedia Security", Vigo, Spain, June 2016 [*DOI :* 10.1145/2909827.2930793], https://hal.archives-ouvertes.fr/hal-01317596.

[12] R. BOIS, V. VUKOTIĆ, R. SICRE, C. RAYMOND, G. GRAVIER, P. SÉBILLOT.*IRISA at TRECVid2016: Crossmodality, Multimodality and Monomodality for Video Hyperlinking*, in "Working Notes of the TRECVid 2016 Workshop", Gaithersburg, United States, 2016, https://hal.archives-ouvertes.fr/hal-01400275.

[13] V. CLAVEAU, E. KIJAK.*Direct vs. indirect evaluation of distributional thesauri*, in "International Conference on Computational Linguistics, COLING", Osaka, Japan, December 2016, https://hal.archives-ouvertes.fr/hal-01394739.

[14] V. CLAVEAU, E. KIJAK.*Distributional Thesauri for Information Retrieval and vice versa*, in "Language and Resource Conference, LREC", Portoroz, Slovenia, May 2016, https://hal.archives-ouvertes.fr/hal-01394770.

[15] G. GRAVIER, M. RAGOT, L. AMSALEG, R. BOIS, G. JADI, E. JAMET, L. MONCEAUX, P. SÉBILLOT.*Shaping-Up Multimedia Analytics: Needs and Expectations of Media Professionals*, in "22nd International Conference on Multimedia Modelling, Special Session Perspectives on Multimedia Analytics", Miami, United States, January 2016, https://hal.inria.fr/hal-01214829.

[16] A. ISCEN, L. AMSALEG, T. FURON.*Scaling Group Testing Similarity Search*, in "ACM International Conference on Multimedia Retrieval 2016", New York, United States, June 2016 [*DOI :* 10.1145/2911996.2912010], https://hal.inria.fr/hal-01304287.

[17] A. ISCEN, T. FURON.*Group Testing for Identification with Privacy*, in "ACM Workshop on Information Hiding and Multimedia Security", Vigo, Spain, June 2016, https://hal.inria.fr/hal-01309032.

[18] A. ISCEN, M. RABBAT, T. FURON.*Efficient Large-Scale Similarity Search Using Matrix Factorization*, in "2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)", Las Vegas, United States, June 2016, https://hal.inria.fr/hal-01294736.

[19] G. JADI, V. CLAVEAU, B. DAILLE, L. MONCEAUX-CACHARD.*Evaluating Lexical Similarity to build Sentiment Similarity*, in "Language and Resource Conference, LREC", portoroz, Slovenia, May 2016, https://hal.archives-ouvertes.fr/hal-01394768.

[20] G. JADI, L. MONCEAUX, V. CLAVEAU, B. DAILLE.*Opinion Target Expression extraction : from English to French* , in "Traitement Automatique des Langues Naturelles, JEP-TALN-RECITAL", Paris, France, July 2016, https://hal.archives-ouvertes.fr/hal-01397188.

[21] B. Þ. JÓNSSON, L. AMSALEG, H. LEJSEK.*SSD Technology Enables Dynamic Maintenance of Persistent High-Dimensional Indexes*, in "ACM International Conference on Multimedia Retrieval 2016", New York, United States, June 2016 [*DOI :* 10.1145/2911996.2912065], https://hal.inria.fr/hal-01305651.

[22] B. Þ. JÓNSSON, M. WORRING, J. ZAHÁLKA, S. RUDINAC, L. AMSALEG.*Ten Research Questions for Scalable Multimedia Analytics*, in "22nd International Conference on Multimedia Modelling, Special Session Perspectives on Multimedia Analytics", Miami, United States, January 2016, https://hal.inria.fr/hal-01214847.

[23] A. LE GUENNEC, S. MALINOWSKI, R. TAVENARD.*Data Augmentation for Time Series Classification using Convolutional Neural Networks*, in "ECML/PKDD Workshop on Advanced Analytics and Learning on Temporal Data", Riva Del Garda, Italy, September 2016, https://halshs.archives-ouvertes.fr/halshs-01357973.

[24] C. MAIGROT, V. CLAVEAU, E. KIJAK, R. SICRE.*MediaEval 2016: A multimodal system for the Verifying Multimedia Use task*, in "MediaEval 2016: "Verfiying Multimedia Use" task", Hilversum, Netherlands, October 2016 [*DOI :* 10.1145/1235], https://hal.archives-ouvertes.fr/hal-01394785.

[25] C. MAIGROT, E. KIJAK, V. CLAVEAU.*Médias traditionnels, médias sociaux : caractériser la réinformation*, in "TALN 2016 - 23ème Conférence sur le Traitement Automatique des Langues Naturelles", Paris, France, July 2016, https://hal.inria.fr/hal-01349871.

[26] H. MEZAACHE, H. BOUZGOU, C. RAYMOND.*Kernel Principal Components Analysis with Extreme Learning Machines for Wind Speed Prediction*, in "Seventh International Renewable Energy Congress, IREC2016", Hammamet, Tunisia, March 2016, https://hal.inria.fr/hal-01394000.

[27] A. ROYER, G. GRAVIER, V. CLAVEAU.*Audio word similarity for clustering with zero resources based on iterative HMM classification*, in "International Conference on Acoustics, Speech and Signal Processing, ICASSP", Shanghai, China, March 2016, p. 5340 - 5344 [*DOI :* 10.1109/ICASSP.2016.7472697], https://hal.archives-ouvertes.fr/hal-01394757.

[28] G. SARGENT, G. BARBOSA DE FONSECA, I. LYON FREIRE, R. SICRE, Z. K. G. DO PATROCÍNIO JR., S. J. GUIMARÃES, G. GRAVIER.*PUC Minas and IRISA at Multimodal Person Discovery*, in "Working Notes Proceedings of the MediaEval Workshop", Hilversum, Netherlands, CEUR-WS.org, October 2016, vol. 1739, https://hal.archives-ouvertes.fr/hal-01400261.

[29] R. TAVENARD, S. MALINOWSKI.*Cost-Aware Early Classification of Time Series*, in "European Conference on Machine Learning and Principles and Practice of Knowledge Discovery", Riva del Garda, Italy, September 2016, https://halshs.archives-ouvertes.fr/halshs-01339007.

[30] V. VUKOTIC, C. RAYMOND, G. GRAVIER.*A step beyond local observations with a dialog aware bidirectional GRU network for Spoken Language Understanding*, in "Interspeech", San Francisco, United States, September 2016, https://hal.inria.fr/hal-01351733.

[31] V. VUKOTIC, C. RAYMOND, G. GRAVIER.*Bidirectional Joint Representation Learning with Symmetrical Deep Neural Networks for Multimodal and Crossmodal Applications*, in "ICMR", New York, United States, ACM, June 2016, https://hal.inria.fr/hal-01314302.

[32] V. VUKOTIĆ, C. RAYMOND, G. GRAVIER.*Multimodal and Crossmodal Representation Learning from Textual and Visual Features with Bidirectional Deep Neural Networks for Video Hyperlinking*, in "ACM Multimedia 2016 Workshop: Vision and Language Integration Meets Multimedia Fusion (iV&L-MM'16)", Amsterdam, Netherlands, ACM Multimedia, October 2016, https://hal.inria.fr/hal-01374727.

[33] H. B. DA SILVA, Z. PATROCINO JR., G. GRAVIER, L. AMSALEG, A. DE A. ARAÚJO, S. J. F. GUIMARÃES.*Near-Duplicate Video Detection Based on an Approximate Similarity Self-Join Strategy*, in "14th International Workshop on Content-based Multimedia Indexing", bucarest, Romania, June 2016, https://hal.inria.fr/hal-01305691.

National Conferences with Proceeding

[34] V. CLAVEAU.*Dimensionnalité intrinsèque dans les espaces de représentation des termes et des documents*, in "Conférence en Recherche d'Information et Applications, CORIA", Toulouse, France, March 2016, https://hal.archives-ouvertes.fr/hal-01394749.

[35] C. MAIGROT, S. BRINGAY, J. AZÉ.*Concept drift vs suicide: comment l'un peut prévenir l'autre ?*, in "EGC: Extraction et Gestion des Connaissances", Reims, France, January 2016, https://hal.inria.fr/hal-01283889.

[36] R. QADER, G. LECORVÉ, D. LOLIVE, P. SÉBILLOT.*Adaptation de la prononciation pour la synthèse de la parole spontanée en utilisant des informations linguistiques*, in "Journées d'Études sur la Parole", Paris, France, July 2016, https://hal.inria.fr/hal-01321361.

[37] A. SIMON, G. GRAVIER, P. SÉBILLOT.*Évaluation d'une nouvelle structuration thématique hiérarchique des textes dans un cadre de résumé automatique et de détection d'ancres au sein de vidéos*, in "Conférence sur le Traitement Automatique des Langues Naturelles", Paris, France, 2016, p. 139-152, https://hal.archives-ouvertes.fr/hal-01399670.

Scientific Books (or Scientific Book chapters)

[38] A. BAILLY, S. MALINOWSKI, R. TAVENARD, L. CHAPEL, T. GUYET.*Dense Bag-of-Temporal-SIFT-Words for Time Series Classification*, in "Advanced Analysis and Learning on Temporal Data", Springer, 2016, https://hal.archives-ouvertes.fr/hal-01252726.

[39] P. BAS, T. FURON, F. CAYRE, G. DOËRR, B. MATHON. , W.-S. GAN, C.-C. J. KUO, T. F. ZHENG, M. BARNI (editors)*Watermarking Security: Fundamentals, Secure Designs and Attacks*, Springer Briefs, springer, January 2016, https://hal.archives-ouvertes.fr/hal-01283066.

[40] T. FURON.*Watermarking security*, in "Information hiding", Information security and privacy series, Artech House, 2016, https://hal.archives-ouvertes.fr/hal-01275880.

Books or Proceedings Editing

[41] L. AMSALEG, G. Þ. GUÐMUNDSSON, C. GURRIN, B. Þ. JÓNSSON, S. SATOH (editors). *Proceedings of the 23rd International Conference on Multimedia Modeling, MMM 2017, Reykjavik, Iceland, January 4-6, 2017*, MultiMedia Modeling, Springer International Publishing, Reykjavík, Iceland, January 2017, vol. 10132, n° 0302-9743 [*DOI :* 10.1007/978-3-319-51811-4], https://hal.archives-ouvertes.fr/hal-01430555.

[42] L. AMSALEG, M. E. HOULE, E. SCHUBERT (editors). *Similarity Search and Applications: 9th International Conference, SISAP 2016, Tokyo, Japan, October 24-26, 2016, Proceedings*, Springer International Publishing, Tokyo, Japan, October 2016, vol. 9939, XVII, 339 [*DOI :* 10.1007/978-3-319-46759-7], https://hal.inria.fr/hal-01391426.

[43] V. CLAVEAU, J. NIE (editors). *Recherche d'information et traitement automatique des langues*, Traitement Automatique des Langues, TAL, ATALA, August 2016, vol. 56, n° 3, https://hal.archives-ouvertes.fr/hal-01394788.

[44] G. GRAVIER, D. CLAIRE-HÉLÈNE, H. BREDIN, I. BOGDAN, B. CHRISTINA, E. DELLANDRÉA, C. JAEYONG, M. A. RIEGLER, S. RICHARD, I. SZÖKE, G. J. F. JONES, M. LARSON (editors). *Working Notes Proceedings of the MediaEval Workshop*, 2016, https://hal.archives-ouvertes.fr/hal-01399674.

Research Reports

[45] G. Þ. GUÐMUNDSSON, L. AMSALEG, B. T. JÓNSSON, M. J. FRANKLIN.*Towards Engineering a Web-Scale Multimedia Service: A Case Study Using Spark*, Inria Rennes Bretagne Atlantique ; Reykjavik University ; UC Berkeley, December 2016, https://hal.inria.fr/hal-01416089.

[46] G. SARGENT, F. BIMBOT, E. VINCENT.*Supplementary material to the article: Estimating the structural segmentation of popular music pieces under regularity constraints*, IRISA-Inria, Campus de Beaulieu, 35042 Rennes cedex ; Inria Nancy, équipe Multispeech, September 2016, https://hal.inria.fr/hal-01368683.

References in notes

[47] M. AGOSTI, J. ALLAN.*Special issue on methods and tools for the automatic construction of hypertext*, in "Information Processing and Retrieval", 1997, vol. 33, n^o 2.

[48] C. A. BHATT, M. S. KANKANHALLI.*Multimedia data mining: State of the art and challenges*, in "Multimedia Tools and Applications", 2011, vol. 51, no 1, p. 35–76.

[49] R. BOIS, A. SIMON, R. SICRE, G. GRAVIER, P. SÉBILLOT, E. MORIN.*IRISA at TrecVid2015: Leveraging Multimodal LDA for Video Hyperlinking*, in "Proc. of the TRECVid Workshop", 2015.

[50] O. CHUM, J. MATAS.*Large-Scale Discovery of Spatially Related Images*, in "Pattern Analysis and Machine Intelligence", 2010, vol. 32, no 2, p. 371-377.

[51] M. ESKEVICH, G. J. JONES, S. CHEN, R. ALY, R. ORDELMAN, D. NADEEM, C. GUINAUDEAU, G. GRAVIER, P. SÉBILLOT, T. DE NIES, P. DEBEVERE, R. V. DE WALLE, P. GALUŠCKOVÁ, P. PECINA, M. LARSON.*Multimedia Information Seeking through Search And Hyperlinking*, in "Intl. Conf. on Multimedia Retrieval", 2013.

[52] C. HERLEY.*ARGOS: Automatically extracting Repeating Objects from multimedia Streams*, in "IEEE Trans. on Multimedia", Feb. 2006, vol. 8, no 1, p. 115–129.

[53] A. PARK, J. R. GLASS.*Unsupervised pattern discovery in speech*, in "IEEE Transaction on Acoustic, Speech and Language Processing", Jan. 2008, vol. 16, no 1, p. 186–197.

[54] K. M. PUA, J. GAUCH, S. E. GUCH, J. Z. MIADOWICZ.*Real time repeated video sequence identification*, in "Computer Vision and Image Understanding", 2004, vol. 93, no 2004, p. 310–327.

[55] A. SIMON, P. SÉBILLOT, G. GRAVIER.*Hierarchical topic structuring: from dense segmentation to topically focused fragments via burst analysis*, in "Recent Advances on Natural Language Processing", Hissar, Bulgaria, 2015, https://hal.archives-ouvertes.fr/hal-01186443.

[56] C. G. M. SNOEK, M. WORRING.*Multimodal Video Indexing: A Review of the State-of-the-Art*, in "Multimedia Tools and Applications", 2005, vol. 25, no 1, p. 5–35.

[57] O. DE ROOIJ, M. WORRING.*Browsing Video Along Multiple Threads*, in "IEEE Transactions on Multimedia", 2010, vol. 12, no 2, p. 121–130.

Project-Team MIMETIC

Analysis-Synthesis Approach for Virtual Human Simulation

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

IN PARTNERSHIP WITH:
Université Haute Bretagne (Rennes 2)

Université Rennes 1

École normale supérieure de Rennes

RESEARCH CENTER
Rennes - Bretagne-Atlantique

THEME
Interaction and visualization

Table of contents

Project-Team MIMETIC

Creation of the Team: 2011 January 01, updated into Project-Team: 2014 January 01

Keywords:

Computer Science and Digital Science:

5.1.5. - Body-based interfaces
5.4.2. - Activity recognition
5.5.4. - Animation
5.6. - Virtual reality, augmented reality

Other Research Topics and Application Domains:

1.3.2. - Cognitive science
2.5. - Handicap and personal assistances
2.8. - Sports, performance, motor skills
5.1. - Factory of the future
5.8. - Learning and training
7.1.1. - Pedestrian traffic and crowds
9.2.2. - Cinema, Television
9.2.3. - Video games
9.3. - Sports

1. Members

Research Scientists

Ludovic Hoyet [Inria, Researcher]
Julien Pettré [Inria, Researcher, in MimeTIC until Feb 2016, HDR]

Faculty Members

Franck Multon [Team leader, Univ. Rennes II, Professor, HDR]
Benoit Bideau [Univ. Rennes II, Professor, HDR]
Nicolas Bideau [Univ. Rennes II, Associate Professor]
Marc Christie [Univ. Rennes I, Associate Professor]
Armel Crétual [Univ. Rennes II, Associate Professor, HDR]
Georges Dumont [ENS Rennes, Professor, HDR]
Richard Kulpa [Univ. Rennes II, Associate Professor, HDR]
Fabrice Lamarche [Univ. Rennes I, Associate Professor]
Antonio Mucherino [Univ. Rennes I, Associate Professor, in MimeTIC since Sept 2016]
Guillaume Nicolas [Univ. Rennes II, Associate Professor]
Anne-Hélène Olivier [Univ. Rennes II, Associate Professor]
Charles Pontonnier [ENS Rennes, Associate Professor]
Camille Pouliquen [Univ. Rennes II, Associate Professor, since Sept. 2016]
Hui-Yin Wu [Univ. Rennes I, Associate Professor, since Sept. 2016]

Technical Staff

Ronan Gaugne [Univ. Rennes I, SED Research Engineer, involved in MimeTIC at 15%]
Christophe Lino [Inria]
Anthony Sorel [Univ. Rennes II]
David Wolinski [Inria, until Nov 2016, granted by ANR PERCOLATION project]

PhD Students

Yacine Boulahia [INSA Rennes]
Julien Bruneau [Univ. Rennes I]
Ana Lucia Cruz Ruiz [Inria, until Nov 2016, granted by ANR ENTRACTE project]
Charles Faure [Univ. Rennes II]
Sean Lynch [Univ. Rennes II]
Marion Morel [Univ. Rennes II, UPMC]
Antoine Muller [ENS Rennes]
Pierre Plantard [Faurecia, until Oct 2016]
Camille Pouliquen [Univ. Rennes II, Until Aug. 2016]
Pierre Touzard [Univ. Rennes II]
Hui-Yin Wu [Univ. Rennes I, until Aug 2016]

Post-Doctoral Fellows

Panayiotis Charalambous [Inria, until May 2016]
Diane Haering [Inria]
Zhiguang Liu [Inria, from Nov 2016]
Laurentius Meerhoff [Inria, from May 2016]

Visiting Scientist

Yijun Shen [Northumbria University, from Sep 2016]

Administrative Assistant

Nathalie Denis [Inria]

2. Overall Objectives

2.1. Presentation

MimeTIC is a multidisciplinary team whose aim is to better understand and model human activity in order to simulate realistic autonomous virtual humans: realistic behaviors, realistic motions and realistic interactions with other characters and users. It leads to modeling the complexity of a human body, as well as of his environment where he can pick-up information and he can act on it. A specific focus is dedicated to human physical activity and sports as it raises the highest constraints and the highest complexity when addressing these problems. Thus, MimeTIC is composed of experts in computer science whose research interests are computer animation, behavioral simulation, motion simulation, crowds and interaction between real and virtual humans. MimeTIC is also composed of experts in sports science, motion analysis, motion sensing, biomechanics and motion control. Hence, the scientific foundations of MimeTIC are motion sciences (biomechanics, motion control, perception-action coupling, motion analysis), computational geometry (modeling of the 3D environment, motion planning, path planning) and design of protocols in immersive environments (use of virtual reality facilities to analyze human activity).

Thanks to these skills, we wish to reach the following objectives: to make virtual humans behave, move and interact in a natural manner in order to increase immersion and to improve knowledge on human motion control. In real situations (see Figure 1), people have to deal with their physiological, biomechanical and neurophysiological capabilities in order to reach a complex goal. Hence MimeTIC addresses the problem of modeling the anatomical, biomechanical and physiological properties of human beings. Moreover these characters have to deal with their environment. Firstly they have to perceive this environment and pick-up relevant information. MimeTIC thus addresses the problem of modeling the environment including its geometry and associated semantic information. Secondly, they have to act on this environment to reach their goals. It leads to cognitive processes, motion planning, joint coordination and force production in order to act on this environment.

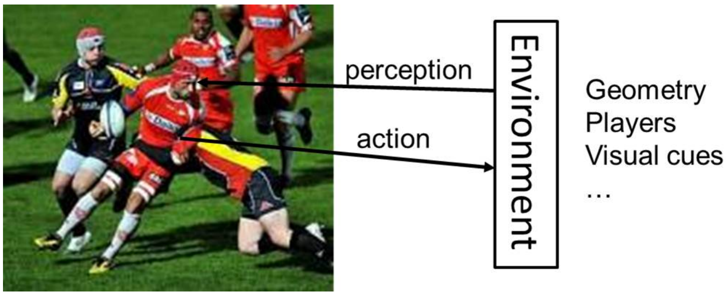


Figure 1. Main objective of MimeTIC: to better understand human activity in order to improve virtual human simulations. It involves modeling the complexity of human bodies, as well as of environments where to pick-up information and act upon.

In order to reach the above objectives, MimeTIC has to address three main challenges:

- dealing with the intrinsic complexity of human beings, especially when addressing the problem of interactions between people for which it is impossible to predict and model all the possible states of the system,
- making the different components of human activity control (such as the biomechanical and physical, the reactive, cognitive, rational and social layers) interact while each of them is modeled with completely different states and time sampling,
- and being able to measure human activity while balancing between ecological and controllable protocols, and to be able to extract relevant information in wide databases of information.

Contrary to many classical approaches in computer simulation, which mostly propose simulation without trying to understand how real people do, the team promotes a coupling between human activity analysis and synthesis, as shown in Figure 2.

In this research path, **improving knowledge on human activity** enables us to highlight fundamental assumptions about natural control of human activities. These contributions can be promoted in e.g. biomechanics, motion sciences, neurosciences. According to these assumptions we propose new algorithms for controlling **autonomous virtual humans**. The virtual humans can perceive their environment and decide of the most natural action to reach a given goal. This work is promoted in computer animation, virtual reality and has some applications in robotics through collaborations. Once autonomous virtual humans have the ability to act as real humans would in the same situation, it is possible to make them **interact with others**, i.e., autonomous characters (for crowds or group simulations) as well as real users. The key idea here is to analyze to what extent the assumptions proposed at the first stage lead to natural interactions with real users. This process enables the validation of both our assumptions and our models.

Among all the problems and challenges described above, MimeTIC focuses on the following domains of research:

- motion sensing which is a key issue to extract information from raw motion capture systems and thus to propose assumptions on how people control their activity,
- human activity & virtual reality, which is explored through sports application in MimeTIC. This domain enables the design of new methods for analyzing the perception-action coupling in human activity, and to validate whether the autonomous characters lead to natural interactions with users,
- interactions in small and large groups of individuals, to understand and model interactions with lot of individual variability such as in crowds,

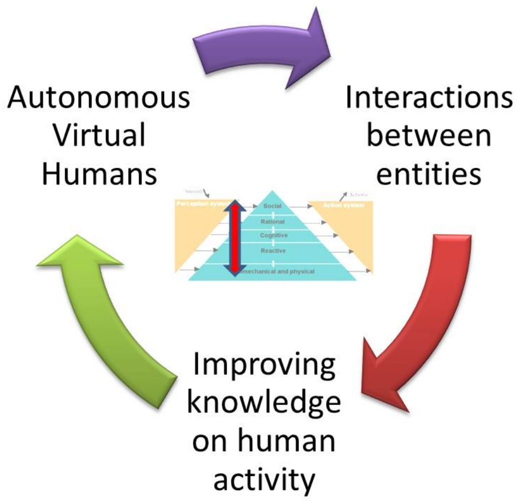


Figure 2. Research path of MimeTIC: coupling analysis and synthesis of human activity enables us to create more realistic autonomous characters and to evaluate assumptions about human motion control.

- virtual storytelling which enables us to design and simulate complex scenarios involving several humans who have to satisfy numerous complex constraints (such as adapting to the real-time environment in order to play an imposed scenario), and to design the coupling with the camera scenario to provide the user with a real cinematographic experience,
- biomechanics which is essential to offer autonomous virtual humans who can react to physical constraints in order to reach high-level goals, such as maintaining balance in dynamic situations or selecting a natural motor behavior among the whole theoretical solution space for a given task,
- and autonomous characters which is a transversal domain that can reuse the results of all the other domains to make these heterogeneous assumptions and models provide the character with natural behaviors and autonomy.

3. Research Program

3.1. Biomechanics and Motion Control

Human motion control is a very complex phenomenon that involves several layered systems, as shown in Figure 3. Each layer of this controller is responsible for dealing with perceptual stimuli in order to decide the actions that should be applied to the human body and his environment. Due to the intrinsic complexity of the information (internal representation of the body and mental state, external representation of the environment) used to perform this task, it is almost impossible to model all the possible states of the system. Even for simple problems, there generally exists an infinity of solutions. For example, from the biomechanical point of view, there are much more actuators (i.e. muscles) than degrees of freedom leading to an infinity of muscle activation patterns for a unique joint rotation. From the reactive point of view there exists an infinity of paths to avoid a given obstacle in navigation tasks. At each layer, the key problem is to understand how people select one solution among these infinite state spaces. Several scientific domains have addressed this problem with specific points of view, such as physiology, biomechanics, neurosciences and psychology.

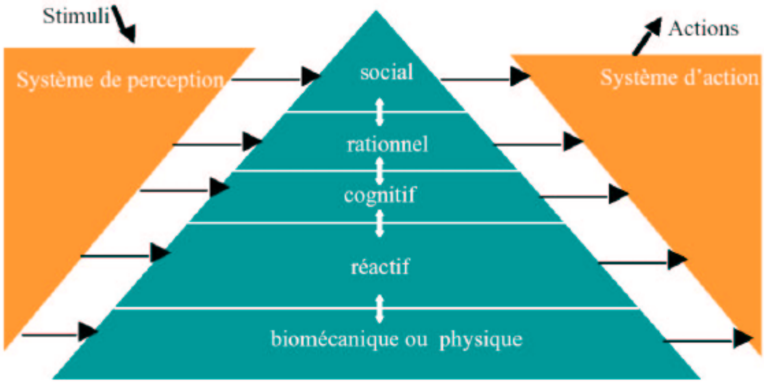


Figure 3. Layers of the motion control natural system in humans.

In biomechanics and physiology, researchers have proposed hypotheses based on accurate joint modeling (to identify the real anatomical rotational axes), energy minimization, force and torques minimization, comfort maximization (i.e. avoiding joint limits), and physiological limitations in muscle force production. All these constraints have been used in optimal controllers to simulate natural motions. The main problem is thus to define how these constraints are composed altogether such as searching the weights used to linearly combine these criteria in order to generate a natural motion. Musculoskeletal models are stereotyped examples for which there exists an infinity of muscle activation patterns, especially when dealing with antagonist muscles. An unresolved problem is to define how to use the above criteria to retrieve the actual activation patterns, while optimization approaches still leads to unrealistic ones. It is still an open problem that will require multidisciplinary skills including computer simulation, constraint solving, biomechanics, optimal control, physiology and neurosciences.

In neurosciences, researchers have proposed other theories, such as coordination patterns between joints driven by simplifications of the variables used to control the motion. The key idea is to assume that instead of controlling all the degrees of freedom, people control higher level variables which correspond to combinations of joint angles. In walking, data reduction techniques such as Principal Component Analysis have shown that lower-limb joint angles are generally projected on a unique plane whose angle in the state space is associated with energy expenditure. Although knowledge exists for specific motions, such as locomotion or grasping, this type of approach is still difficult to generalize. The key problem is that many variables are coupled and it is very difficult to objectively study the behavior of a unique variable in various motor tasks. Computer simulation is a promising method to evaluate such type of assumptions as it enables to accurately control all the variables and to check if it leads to natural movements.

Neurosciences also address the problem of coupling perception and action by providing control laws based on visual cues (or any other senses), such as determining how the optical flow is used to control direction in navigation tasks, while dealing with collision avoidance or interception. Coupling of the control variables is enhanced in this case as the state of the body is enriched by the large amount of external information that the subject can use. Virtual environments inhabited with autonomous characters whose behavior is driven by motion control assumptions is a promising approach to solve this problem. For example, an interesting problem in this field is navigation in an environment inhabited with other people. Typically, avoiding static obstacles together with other people displacing into the environment is a combinatory problem that strongly relies on the coupling between perception and action.

One of the main objectives of MimeTIC is to enhance knowledge on human motion control by developing innovative experiments based on computer simulation and immersive environments. To this end, designing experimental protocols is a key point and some of the researchers in MimeTIC have developed this skill in biomechanics and perception-action coupling. Associating these researchers to experts in virtual human simulation, computational geometry and constraints solving enable us to contribute to enhance fundamental knowledge in human motion control.

3.2. Experiments in Virtual Reality

Understanding interactions between humans is very challenging because it addresses many complex phenomena including perception, decision-making, cognition and social behaviors. Moreover, all these phenomena are difficult to isolate in real situations, and it is therefore very complex to understand their individual influence on these human interactions. It is then necessary to find an alternative solution that can standardize the experiments and that allows the modification of only one parameter at a time. Video was first used since the displayed experiment is perfectly repeatable and cut-offs (stop the video at a specific time before its end) allow having temporal information. Nevertheless, the absence of adapted viewpoint and stereoscopic vision does not provide depth information that are very meaningful. Moreover, during video recording session, the real human is acting in front of a camera and not of an opponent. The interaction is then not a real interaction between humans.

Virtual Reality (VR) systems allow full standardization of the experimental situations and the complete control of the virtual environment. It is then possible to modify only one parameter at a time and to observe its influence on the perception of the immersed subject. VR can then be used to understand what information is picked up to make a decision. Moreover, cut-offs can also be used to obtain temporal information about when information is picked up. When the subject can moreover react as in a real situation, his movement (captured in real time) provides information about his reactions to the modified parameter. Not only is the perception studied, but the complete perception-action loop. Perception and action are indeed coupled and influence each other as suggested by Gibson in 1979.

Finally, VR allows the validation of the virtual human models. Some models are indeed based on the interaction between the virtual character and the other humans, such as a walking model. In that case, there are two ways to validate it. First, they can be compared to real data (e.g. real trajectories of pedestrians). But such data are not always available and are difficult to get. The alternative solution is then to use VR. The validation of the realism of the model is then done by immersing a real subject in a virtual environment in which a virtual character is controlled by the model. Its evaluation is then deduced from how the immersed subject reacts when interacting with the model and how realistic he feels the virtual character is.

3.3. Computational Geometry

Computational geometry is a branch of computer science devoted to the study of algorithms which can be stated in terms of geometry. It aims at studying algorithms for combinatorial, topological and metric problems concerning sets of points in Euclidian spaces. Combinatorial computational geometry focuses on three main problem classes: static problems, geometric query problems and dynamic problems.

In static problems, some inputs are given and the corresponding outputs need to be constructed or found. Such problems include linear programming, Delaunay triangulations, and Euclidian shortest paths for instance. In geometric query problems, commonly known as geometric search problems, the input consists of two parts: the search space part and the query part, which varies over the problem instances. The search space typically needs to be preprocessed, in a way that multiple queries can be answered efficiently. Some typical problems are range searching, point location in a portioned space, or nearest neighbor queries. In dynamic problems, the goal is to find an efficient algorithm for finding a solution repeatedly after each incremental modification of the input data (addition, deletion or motion of input geometric elements). Algorithms for problems of this type typically involve dynamic data structures. Both of previous problem types can be converted into a dynamic problem, for instance, maintaining a Delaunay triangulation between moving points.

In this context, distance geometry relies solely on distances, instead of points and lines, as in classical geometry. Various applications lead to the definition of problems that can be formulated as a distance geometry, including sensor network localization, robot coordination, the identification of molecular conformations, or as in the context of MimeTIC relations between objects in virtual scenes (e.g., distances between body segments, agents, or cameras). In recent years, scientific research has been oriented to the assumptions allowing for discretizing the search space of a given distance geometry problem. The discretization (which is exact in some situations) allows to conceive ad-hoc and efficient algorithms, and for enumerating the entire solution set of a given instance.

The Mimetic team works on problems such as crowd simulation, spatial analysis, path and motion planning in static and dynamic environments, camera planning with visibility constraints for instance. The core of those problems, by nature, relies on problems and techniques belonging to computational geometry. Proposed models pay attention to algorithms complexity to be compatible with performance constraints imposed by interactive applications.

4. Application Domains

4.1. Autonomous Characters

Autonomous characters are becoming more and more popular as they are used in an increasing number of application domains. In the field of special effects, virtual characters are used to replace secondary actors and generate highly populated scenes that would be hard and costly to produce with real actors. In video games and virtual storytelling, autonomous characters play the role of actors that are driven by a scenario. Their autonomy allows them to react to unpredictable user interactions and adapt their behavior accordingly. In the field of simulation, autonomous characters are used to simulate the behavior of humans in different kind of situations. They enable to study new situations and their possible outcomes.

One of the main challenges in the field of autonomous characters is to provide a unified architecture for the modeling of their behavior. This architecture includes perception, action and decisional parts. This decisional part needs to mix different kinds of models, acting at different time scale and working with different nature of data, ranging from numerical (motion control, reactive behaviors) to symbolic (goal oriented behaviors, reasoning about actions and changes).

In the MimeTIC team, we focus on autonomous virtual humans. Our problem is not to reproduce the human intelligence but to propose an architecture making it possible to model credible behaviors of anthropomorphic virtual actors evolving/moving in real time in virtual worlds. The latter can represent particular situations studied by psychologists of the behavior or to correspond to an imaginary universe described by a scenario writer. The proposed architecture should mimic all the human intellectual and physical functions.

4.2. Biomechanics and Motion Analysis

Biomechanics is obviously a very large domain. This large set can be divided regarding to the scale at which the analysis is performed going from microscopic evaluation of biological tissues' mechanical properties to macroscopic analysis and modeling of whole body motion. Our topics in the domain of biomechanics mainly lie within this last scope. In order to obtain a better understanding of human motion, MimeTIC addresses three main situations: everyday motions of a lambda subject, locomotion of pathological subjects and sports gestures.

In the first situation, MimeTIC is interested in studying how subjects maintain their balance in highly dynamic conditions. Until now, balance have nearly always been considered in static or quasi-static conditions. The knowledge of much more dynamic cases still has to be improved. Our approach has demonstrated that, first of all, the question of the parameter that will allow to do this is still open. We have also largely contributed to gaining a better understanding of collision avoidance between pedestrians. This topic includes the research of the parameters that are interactively controlled and the study of each one's role within this interaction.

The second situation focuses on locomotion of pathological subjects. When patients cannot walk efficiently, in particular those suffering from central nervous system affections, it becomes very useful for practitioners to benefit from an objective evaluation of their capacities. To facilitate such evaluations, we have developed two complementary indices, one based on kinematics and the other one on muscle activations. One major point of our research is that such indices are usually only developed for children whereas adults with these affections are much more numerous.

Finally, in sports, where gesture can be considered, in some way, as abnormal, the goal is more precisely to understand the determinants of performance. This could then be used to improve training programs or devices. Two different sports have been studied: a) the tennis serve, where the goal was to understand the contribution of each segment of the body on the speed of the ball and b) the influence of the mechanical characteristics of the fin in fin swimming.

After having improved the knowledge of these different gestures a second goal is then to propose modeling solutions that can be used in VR environments for other research topics within MimeTIC. This has been the case, for example, for collision avoidance.

4.3. Interactions between walkers

Modeling and simulating the interactions between walkers is a very active, complex and competitive domain, interesting various disciplines such as Mathematics, Cognitive Sciences, Physics, Computer Graphics, etc. Interactions between walkers are by definition at the very core of our society since they represent the basic synergies of our daily life. When walking in the street, we take information about our surrounding environment in order to interact with people, move without collision, alone or in a group, intercept, meet or escape to somebody. Large groups of walkers can be first seen as a complex system: numerous local interactions occur between its elements and result into macroscopic emergent phenomena. Interactions are of various nature (e.g., collision avoidance, following) and are undergoing various factors as well. Physical factors are crucial as a group gathers by definition numerous moving people with a certain level of density. But sociological, cultural and psychological factors are important as well, since people's behavior is deeply changed from country to country, or depending on the considered situations. On the computational point of view, simulating the movements of large groups of walkers (i.e., crowds) pushes traditional simulation algorithms to their limit. As an element of a crowd is subject to interact with any other element belonging the same crowd, a naïve simulation algorithm has a quadratic complexity. Specific strategies are set to face such a difficulty: level-of-detail techniques enable scaling large crowd simulation and reach real-time solutions.

MimeTIC is an international key contributor in the domain of understanding and simulating interactions between walkers, in particular for virtual crowds. Our approach is specific and based on three axes. First, our modeling approach is based on human movement science: we conduct challenging experiments focusing on the perception as well as on the motion involved in local interactions between walkers both using real and virtual set-ups. Second: we develop high-performance solutions for crowd simulation. Third, we develop solutions for realistic navigation in virtual world to enable interaction with crowds in Virtual Reality.

4.4. Motion Sensing of Human Activity

Recording human activity is a key point of many applications and fundamental works. Numerous sensors and systems have been proposed to measure positions, angles or accelerations of the user's body parts. Whatever the system is, one of the main problems is to be able to automatically recognize and analyze the user's performance according to poor and noisy signals. Human activity and motion are subject to variability: intra-variability due to space and time variations of a given motion, but also inter-variability due to different styles and anthropometric dimensions. MimeTIC has addressed the above problems in two main directions.

Firstly, we have studied how to recognize and quantify motions performed by a user when using accurate systems such as Vicon (product of Oxford Metrics) or Optitrack (product of Natural Point) motion capture systems. These systems provide large vectors of accurate information. Due to the size of the state vector (all the degrees of freedom) the challenge is to find the compact information (named features) that enables the automatic system to recognize the performance of the user. Whatever the method used, finding these relevant features that are not sensitive to intra-individual and inter-individual variability is a challenge. Some researchers have proposed to manually edit these features (such as a Boolean value stating if the arm is moving forward or backward) so that the expertise of the designer is directly linked with the success ratio. Many proposals for generic features have been proposed, such as using Laban notation which was introduced to encode dancing motions. Other approaches tend to use machine learning to automatically extract these features. However most of the proposed approaches were used to seek a database for motions which properties correspond to the features of the user's performance (named motion retrieval approaches). This does not ensure the retrieval of the exact performance of the user but a set of motions with similar properties.

Secondly, we wish to find alternatives to the above approach which is based on analyzing accurate and complete knowledge on joint angles and positions. Hence new sensors, such as depth-cameras (Kinect, product of Microsoft) provide us with very noisy joint information but also with the surface of the user. Classical approaches would try to fit a skeleton into the surface in order to compute joint angles which, again, lead to large state vectors. An alternative would be to extract relevant information directly from the raw data, such as the surface provided by depth cameras. The key problem is that the nature of these data may be very different from classical representation of human performance. In MimeTIC, we try to address this problem in specific application domains that require picking specific information, such as gait asymmetry or regularity for clinical analysis of human walking.

4.5. VR and Sports

Sport is characterized by complex displacements and motions. These motions are dependent on visual information that the athlete can pick up in his environment, including the opponent's actions. Perception is thus fundamental to the performance. Indeed, a sportive action, as unique, complex and often limited in time, requires a selective gathering of information. This perception is often seen as a prerogative for action, it then takes the role of a passive collector of information. However, as mentioned by Gibson in 1979, the perception-action relationship should not be considered sequential but rather as a coupling: we perceive to act but we must act to perceive. There would thus be laws of coupling between the informational variables available in the environment and the motor responses of a subject. In other words, athletes have the ability to directly perceive the opportunities of action directly from the environment. Whichever school of thought considered, VR offers new perspectives to address these concepts by complementary using real time motion capture of the immersed athlete.

In addition to better understanding sports and interactions between athletes, VR can also be used as a training environment as it can provide complementary tools to coaches. It is indeed possible to add visual or auditory information to better train an athlete. The knowledge found in perceptual experiments can be for example used to highlight the body parts that are important to look at to correctly anticipate the opponent's action.

4.6. Interactive Digital Storytelling

Interactive digital storytelling, including novel forms of edutainment and serious games, provides access to social and human themes through stories which can take various forms and contains opportunities for massively enhancing the possibilities of interactive entertainment, computer games and digital applications. It provides chances for redefining the experience of narrative through interactive simulations of computer-generated story worlds and opens many challenging questions at the overlap between computational narratives, autonomous behaviours, interactive control, content generation and authoring tools.

Of particular interest for the MimeTIC research team, virtual storytelling triggers challenging opportunities in providing effective models for enforcing autonomous behaviours for characters in complex 3D environments. Offering both low-level capacities to characters such as perceiving the environments, interacting with the environment and reacting to changes in the topology, on which to build higher-levels such as modelling abstract representations for efficient reasoning, planning paths and activities, modelling cognitive states and behaviours requires the provision of expressive, multi-level and efficient computational models. Furthermore virtual storytelling requires the seamless control of the balance between the autonomy of characters and the unfolding of the story through the narrative discourse. Virtual storytelling also raises challenging questions on the conveyance of a narrative through interactive or automated control of the cinematography (how to stage the characters, the lights and the cameras). For example, estimating visibility of key subjects, or performing motion planning for cameras and lights are central issues for which have not received satisfactory answers in the literature.

4.7. VR and Ergonomics

The design of workstations nowadays tends to include assessment steps in a Virtual Environment (VE) to evaluate ergonomic features. This approach is more cost-effective and convenient since working directly on the Digital Mock-Up (DMU) in a VE is preferable to constructing a real physical mock-up in a Real Environment (RE). This is substantiated by the fact that a Virtual Reality (VR) set-up can be easily modified, enabling quick adjustments of the workstation design. Indeed, the aim of integrating ergonomics evaluation tools in VEs is to facilitate the design process, enhance the design efficiency, and reduce the costs.

The development of such platforms asks for several improvements in the field of motion analysis and VR. First, interactions have to be as natural as possible to properly mimic the motions performed in real environments. Second, the fidelity of the simulator also needs to be correctly evaluated. Finally, motion analysis tools have to be able to provide in real-time biomechanics quantities usable by ergonomists to analyse and improve the working conditions.

5. Highlights of the Year

5.1. Highlights of the Year

This year, we feature four of the team's research results as specific highlights, in particular due to their high publication impacts.

Our work entitled "Validation of an ergonomic assessment method using Kinect data in real workplace conditions" ([15] by Pierre Plantard, Hubert PH Shum, Anne-Sophie Le Pierres and Franck Multon) has been accepted in the journal Applied Ergonomics. This publication is very important for future works in ergonomics as it demonstrates the relevance of the Kinect data correction for in-site (on a real workstation in factories) in an ergonomic purpose.

A State of the art paper, "Muscle-Based Control For Character Animation" has been published in Computer Graphics Forum ([6] by Ana Lucia Cruz Ruiz, Charles Pontonnier, Nicolas Pronost and Georges Dumont). It presents an organized review of over a decade of research in muscle-based control for character animation, its fundamental concepts and future directions for development. The core of this review contains a classification of control methods, tables summarizing their key aspects, and popular neuromuscular functions used within these controllers.

Our work entitled "Perceptual Effect of Shoulder Motions on Crowd Animations" ([11] by Ludovic Hoyet, Anne-Hélène Olivier, Richard Kulpa and Julien Pettré) has been accepted and presented in SIGGRAPH 2016, the premier and most selective computer graphics scientific event, and published in ACM Transaction on Graphics. It explores how local interactions between walkers are perceived by users when secondary shoulder motions are displayed, and demonstrates the benefits of such secondary animations in large-scale crowd scenarios.

Two papers exploring the effects of the avatar's representation on users' sense of "virtual embodiment" (i.e., the extent to which we accept an avatar to be our representation in the virtual environment) were published in Frontiers in Robotics and AI [10] and in IEEE VR [19], resulting from a collaboration between Ludovic Hoyet (MimeTIC), and Ferran Argelaguet and Anatole Lécuyer (Hybrid). This work paves the way to further collaborations on understanding how we accept virtual characters as our own representation in virtual environments.

5.1.1. Awards

This year, the ANR Entracte leaded by CNRS/LAAS received the best price for ANR Project in November 2016 in Paris ("Grand prix du Numérique des 10ans de l'ANR", link).

The ANR Jeune Chercheur project Cinecitta, led by Marc Christie, has also been awarded one of the 10 "iconic" projects (*projets emblématiques*) for the 10 years of the ANR, and will be presented at the 10 years celebration of the ANR in December 2016.

6. New Software and Platforms

6.1. AsymGait

Asymmetry index for clinical gait analysis based on depth images
KEYWORDS: Motion analysis - Kinect - Clinical analysis
SCIENTIFIC DESCRIPTION

The system uses depth images delivered by the Microsoft Kinect to retrieve gait cycles. To this end it is based on analyzing the knee trajectories instead of the feet to obtain more robust gait event detection. Based on these cycles, the system computes a mean gait cycle model to decrease the effect of noise of the system. Asymmetry is then computed at each frame of the gait cycle as the spatial difference between the left and right parts of the body.
FUNCTIONAL DESCRIPTION

AsymGait is a software package that works with Microsoft Kinect data, especially depth images, in order to carry-out clinical gait analysis. First is identifies the main gait events using the depth information (footstrike, toe-off) to isolate gait cycles. Then it computes a continuous asymmetry index within the gait cycle. Asymmetry is viewed as a spatial difference between the two sides of the body.

- Participants: Franck Multon and Edouard Auvinet
- Contact: Franck Multon

6.2. Cinematic Viewpoint Generator

KEYWORDS: Virtual Cinematography - Intelligent Gallery
FUNCTIONAL DESCRIPTION

The software, developed as an API, provides a mean to automatically compute a collection of viewpoints over one or two specified geometric entities, in a given 3D scene, at a given time. These viewpoints satisfy classical cinematographic framing conventions and guidelines including different shot scales (from extreme long shot to extreme close-up), different shot angles (internal, external, parallel, apex), and different screen compositions (thirds, fifths, symmetric of di-symmetric). The viewpoints allow to cover the range of possible framings for the specified entities. The computation of such viewpoints relies on a database of framings that are dynamically adapted to the 3D scene by using a manifold parametric representation and guarantee the visibility of the specified entities. The set of viewpoints is also automatically annotated with cinematographic tags such as shot scales, angles, compositions, relative placement of entities, line of interest.

- Participants: Emmanuel Badier, Christophe Lino and Marc Christie
- Partners: Université d'Udine - Université de Nantes - William Bares
- Contact: Marc Christie

6.3. Directors Lens Motion Builder

KEYWORDS: Previzualisation - Virtual cinematography - 3D animation
FUNCTIONAL DESCRIPTION

Directors Lens Motion Builder is a software plugin for Autodesk's Motion Builder animation tool. This plugin features a novel workflow to rapidly prototype cinematographic sequences in a 3D scene, and is dedicated to the 3D animation and movie previsualization industries. The workflow integrates the automated computation of viewpoints (using the Cinematic Viewpoint Generator) to interactively explore different framings of the scene, proposes means to interactively control framings in the image space, and proposes a technique to automatically retarget a camera trajectory from one scene to another while enforcing visual properties. The tool also proposes to edit the cinematographic sequence and export the animation. The software can be linked to different virtual camera systems available on the market.

- Participants: Emmanuel Badier, Christophe Lino and Marc Christie
- Partner: Université de Rennes 1
- Contact: Marc Christie

6.4. Kimea

Kinect IMprovement for Egronomics Assessment
KEYWORDS: Biomechanics - Motion analysis - Kinect
SCIENTIFIC DESCRIPTION

Kimea consists in correcting skeleton data delivered by a Microsoft Kinect for ergonomics purposes. Kimea is able to manage most of the occultations that can occur on workstations (real working situations). To this end, Kimea relies on a database of examples/poses organized as a graph, in order to replace unreliable body segment reconstructions by poses that have already been measured on real subjects. The potential pose candidates are used in an optimization framework.
FUNCTIONAL DESCRIPTION

Kimea gets Kinect data as input data (skeleton data) and correct most of measurement errors to carry-out ergonomic assessment at workstation.

- Participants: Franck Multon, Pierre Plantard and Hubert Shum
- Partner: Faurecia
- Contact: Franck Multon

6.5. Populate

KEYWORDS: Behavioral animation - Virtual cities
SCIENTIFIC DESCRIPTION

Populate is a toolkit dedicated to task scheduling under time and space constraints in the field of behavioral animation. It is currently used to populate virtual cities with pedestrians performing different kind of activities implying travels between different locations. However the generic aspect of the algorithm and underlying representations enables its use in a wide range of applications that need to link activity, time and space. The main scheduling algorithm relies on the following inputs: an informed environment description, an activity an agent needs to perform and individual characteristics of this agent. The algorithm produces a valid task schedule compatible with time and spatial constraints imposed by the activity description and the environment. In this task schedule, time intervals relating to travel and task fulfilment are identified and locations where tasks should be performed are automatically selected.
With a good configuration of agents characteristics (based on statistics), we demonstrated that tasks schedules produced by Populate are representative of human ones. In conjunction with TopoPlan, it has been used to populate a district of Paris as well as imaginary cities with several thousands of pedestrians navigating in real time.
FUNCTIONAL DESCRIPTION

The software provides the following functionalities:

- A high level XML dialect that is dedicated to the description of agents activities in terms of tasks and sub activities that can be combined with different kind of operators: sequential, without order, interlaced. This dialect also enables the description of time and location constraints associated to tasks.
- An XML dialect that enables the description of agent's personal characteristics.
- An informed graph describes the topology of the environment as well as the locations where tasks can be performed. A bridge between TopoPlan and Populate has also been designed. It provides an automatic analysis of an informed 3D environment that is used to generate an informed graph compatible with Populate.
- The generation of a valid task schedule based on the previously mentioned descriptions.
- Participants: Fabrice Lamarche and Carl-Johan Jorgensen
- Contact: Fabrice Lamarche

6.6. The Theater

KEYWORDS: Interactive Scenarios - 3D animation - Storytelling
SCIENTIFIC DESCRIPTION

The Theater is a software framework to develop interactive scenarios in virtual 3D environments. The framework provides means to author and orchestrate 3D character behaviors and simulate them in real-time. The tools provide a basis to build a range of 3D applications, from simple simulations with reactive behaviors, to complex storytelling applications including narrative mechanisms such as flashbacks.

FUNCTIONAL DESCRIPTION

The Theater is Unity 3D application. XML descriptions are used to specify characters behaviors.

- Contact: Marc Christie

6.7. Immerstar Platform

Participants: Georges Dumont [contact], Ronan Gaugne, Anthony Sorel, Franck Multon.

With the two platforms of virtual reality, Immersia and Immermove, grouped under the name Immerstar, the team has access to high level scientific facilities. This equipment benefits the research teams of the center and has allowed them to extend their local, national and international collaborations. The Immerstar platform is granted by a Inria CPER funding for 2015-2019 that enables important evolutions of the equipment. In 2016, the first technical evolutions have been decided, with, for Immermove, the addition of a third face to the immersive space, and the extension of the Vicon tracking system, and for Immersia, the installation of WQXGA laser projectors and of a new tracking system.

7. New Results

7.1. Outline

In 2016, MimeTIC pursued its efforts in improving virtual human simulation by initiating new projects in this domain, such as the Inria PRE with CAIRN team, and recruiting Antonio Mucherino in Inria half-delegation. Our main goal is to provide more natural human motion in real-time applications, which is a transversal requirement in many of MimeTIC's research domains.

- In Biomechanics, being able to rapidly simulate plausible human motion enables to explore new approaches to provide real-time feedback to users in many application domains, such a rehabilitation, sports training, ergonomics and industrial training.

- In computer graphics, simulating natural motion either relies on heavy mechanical simulation and optimal control or adapting motion capture data. We wish to push dynamic simulation a step forward to propose new biomechanically-based simulation, such as actuating the virtual human with muscles instead of rotating servos. We also wish to simplify the process of retargeting motion capture data, which is a process still difficult to automatize. In both cases, we also promote the idea of understanding how human perception behaves when facing inaccurate simulation, in order to provide accurate simulations only when necessary.
- In virtual reality, real-time motion capture and simulation are essential when using head mounted display devices as users cannot perceive their own body during immersive experiences. Hence, simulating natural avatar motion and reacting efficiently to the user's actions are key points to ensure good Presence and Embodiment. MimeTIC is collaborating with other teams in VR, such as Hybrid, to address this complex pluridisciplinary question.
- In digital storytelling, interactive autonomous virtual characters lever the potentiality of proposing complex stories on social and human themes. More stories are now created with the goal of proposing several interactive storylines, which massively enhances the possibilities of interactive entertainment, computer games and digital applications. Projects in MimeTIC explore for instance how to provide a seamless control of the balance between the autonomy of characters and the unfolding of the story through the narrative discourse.

Hence, the organization of the results is reflecting these main challenges in motion analysis, virtual human simulation, interaction in VR, and digital storytelling.

7.2. Motion Analysis

In motion analysis, we continued designing new approaches to measure human performance in specific applications, such as clinical gait assessment, ergonomics and sports. We also developed an original approach to concurrently analyze and simulate human motion, by addressing the problem of redundancy in musculoskeletal models.

7.2.1. Clinical gait assessment based on Kinect data

Participant: Franck Multon.

In clinical gait analysis, we proposed a method to overcome the main limitations imposed by the low accuracy of the Kinect measurements in real medical exams. Indeed, inaccuracies in the 3D depth images lead to badly reconstructed poses and inaccurate gait event detection. In the latter case, confusion between the foot and the ground leads to inaccuracies in the foot-strike and toe-off event detection, which are essential information to get in a clinical exam. To tackle this problem we assumed that heel strike events could be indirectly estimated by searching for the extreme values of the distance between the knee joints along the walking longitudinal axis. As Kinect sensor may not accurately locate the knee joint, we used anthropometrical data to select a body point located at a constant height where the knee should be in the reference posture. Compared to previous works using a Kinect, heel strike events and gait cycles are more accurately estimated, which could improve global clinical gait analysis frameworks with such a sensor. Once these events are correctly detected, it is possible to define indexes that enable the clinician to have a rapid state of the quality of the gait. We therefore proposed a new method to assess gait asymmetry based on depth images, to decrease the impact of errors in the Kinect joint tracking system. It is based on the longitudinal spatial difference between lower-limb movements during the gait cycle. The movement of artificially impaired gaits was recorded using both a Kinect placed in front of the subject and a motion capture system. The proposed longitudinal index distinguished asymmetrical gait, while other symmetry indices based on spatiotemporal gait parameters failed using such Kinect skeleton measurements. This gait asymmetry index measured with a Kinect is low cost, easy to use and is a promising development for clinical gait analysis.

This method has been challenged with other classical approaches to assess gait asymmetry using either cheap Kinect data or Vicon data. We demonstrate the superiority of the approach when using Kinect data for which traditional approaches failed to accurately detect gait asymmetry. It has been validated on healthy subjects who were forced to walk with a 5cm sole placed below each foot alternatively [2].

This work has been done in collaboration with the MsKLab from Imperial College London, to design new gait asymmetry indexes that could be used in daily clinical analysis.

7.2.2. New automatic methods to assess motion in industrial contexts based on Kinect

Participants: Franck Multon, Pierre Plantard.

Recording human activity is a key point of many applications and fundamental works. Numerous sensors and systems have been proposed to measure positions, angles or accelerations of the user's body parts. Whatever the system is, one of the main challenge is to be able to automatically recognize and analyze the user's performance according to poor and noisy signals. Hence, recognizing and measuring human performance are important scientific challenges especially when using low-cost and noisy motion capture systems. MimeTIC has addressed the above problems in two main application domains. In this section, we detail the ergonomics application of such an approach.

Firstly, in ergonomics, we explored the use of low-cost motion capture systems (i.e., a Microsoft Kinect) to measure the 3D pose of a subject in natural environments, such as on a workstation, with many occlusions and inappropriate sensor placements. Predicting the potential accuracy of the measurement for such complex 3D poses and sensor placements is challenging with classical experimental setups. After having evaluated the actual accuracy of the pose reconstruction method delivered by the Kinect, we have identified that occlusions were a very important problem to solve in order to obtain reliable ergonomic assessments in real cluttered environments. To this end, we extended previous correction methods proposed by Hubert Shum (Northumbria University) which consist in identifying the reliable and unreliable parts of the Kinect skeleton data, and to replace unreliable ones by prior knowledge recorded in a database. In collaboration with Hubert Shum, we extended this approach to deal with long occlusions that occur in real manufacturing conditions. To this end we proposed a new data structure named Filtered Pose Graph to speed-up the process, and select example poses that improve the quality of the correction, especially ensuring continuity. We have demonstrated a significant increase of the quality of the correction, especially when large tracking errors occur with the Kinect system [16].

This method has been applied to a complete ergonomic process outputting RULA scores based on the reconstructed and corrected poses. We also demonstrated that it delivers new ergonomic information compared to traditional approaches based on isolated pictures: it provides time spent above a given RULA score which is a valuable information to support decision in ergonomics [15]. We also challenged this method with a reference motion capture system in laboratory conditions. In order to evaluate the actual use in ergonomics, we also compared the ergonomic scores obtained with this automatic method to two experts' scores in real factories. The results show very good agreements between automatic and manual assessments, and have been published in Applied Ergonomics journal [25].

This work was partially funded by the Faurecia company through a Cifre convention.

7.2.3. Evaluation and analysis of sports gestures: application to tennis serve

Participants: Richard Kulpa, Marion Morel, Benoit Bideau, Pierre Touzard.

Following the previous studies we made on tennis serve, we were able to evaluate the link between performance and risk of injuries. To go further, we made new experiments to quantify the influence of fatigue on the performance of tennis serve, that is to say the kinematic, kinetic and performance changes that occur in the serve throughout a prolonged tennis match play [12], [13]. To this end, we recorded serves of several advanced tennis players with a motion capture system before, at mid-match, and after a 3-hour tennis match. Before and after each match, we also recorded electromyographic data of 8 upper limb muscles obtained during isometric maximal voluntary contraction. These experiments showed a decrease in mean power frequency values for several upper limb muscles that is an indicator of local muscular fatigue. Decreases in serve ball speed, ball impact height, maximal angular velocities and an increase in rating of perceived exertion were also observed between beginning and end of match. However, no change in timing of maximal angular velocities was observed. The consistency in timing of maximal angular velocities suggests that advanced tennis players are able to maintain the temporal pattern of their serve technique, in spite of the muscular fatigue development [12].

Moreover, we showed that passive shoulder internal rotation and total range of motion are significantly decreased during a 3-hour tennis match that is identified as an injury risk factor among tennis players [13].

Overall, automatically evaluating and quantifying the performance of a player is a complex task since the important motion features to analyze depend on the type of performed action. But above all, this complexity is due to the variability of morphologies and styles of both novices and experts (who perform the reference motions). Only based on a database of experts' motions and no additional knowledge, we propose an innovative 2-level DTW (Dynamic Time Warping) approach to temporally and spatially align the motions and extract the imperfections of the novice's performance for each joint. We applied our method on tennis serves and karate katas [22].

7.2.4. Interactions between walkers

Participants: Anne-Hélène Olivier, Armel Crétual, Julien Bruneau, Richard Kulpa, Sean Lynch, Laurentius Meerhoff, Julien Pettré.

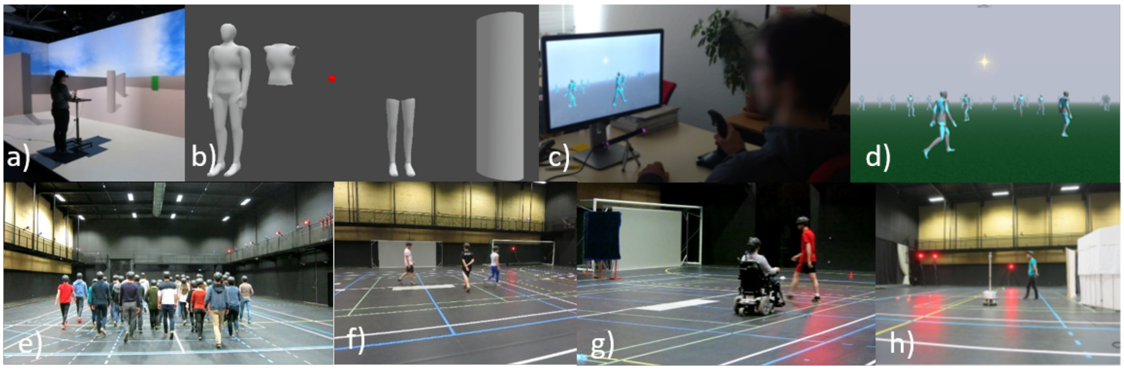


Figure 4. Experiments performed to investigate interactions between walkers.

Interaction between people, and especially local interaction between walkers, is a main research topic of MimeTIC. We propose experimental approaches using both real and virtual environments to study both perception and action aspects of the interaction. This year, we developed new experiments in our immersive platform. In the context of Sean Lynch's PhD on the visual perception of human motion during interactions in locomotor tasks, we designed a study to investigate whether local limb motion is required to successfully avoid a single dynamic obstacle or if global motion alone provides sufficient information (Figures 4.a and 4.b). Sixteen healthy subjects were immersed in a virtual environment that required navigating towards a target, whilst an obstacle crossed its path. Within the virtual environment, four occluding walls prevented the subject observing the complete environment at the initiation of movement, ensuring steady state was reached prior to obstacle interaction. The velocity and heading of the obstacle were programmed to result in a range of future crossing distance (varying from 0.1 to 1.2m) in front and behind the subject. The velocity and heading of the obstacle were fixed, and the subject used a joystick to control its orientation to avoid collision. Five obstacle appearances were presented in a randomized order; a full body (control condition), trunk- or legs-only (i.e., local motion only), and a cylinder or sphere representing the center of gravity (COG) (i.e., global motion only). No significant difference for obstacle appearance was found on number of collisions. However, in both global motion only conditions, subjects adopted alternative collision avoidance strategies compared to the full body control condition. Distance regulation and collision avoidance within daily activities may be principally regulated by global rather than local motion. Underlying mechanisms may differ accordingly to shape and size, however there is no impediment for successful completion of collision avoidance.

Second, we provide lot of efforts to investigate the complex case of multiple interactions while in previous studies we mainly focused on pairwise interactions. We developed a new experiment using an eye tracker to provide insight about the selection process of the interactions (Figures 4.c and 4.d). We proposed to study the human gaze during a navigation task in a crowded virtual environment. The characteristics of each virtual agent was known and controlled. Then, by recording the gaze activity, we are able to highlight the characteristics of each agent the participant was looking at. Results first showed a strong link between the fixated agents and the trajectory adaptations of the participants which means that participants looked at agents they are interacting with, which is an important result to validate the use of the eye tracker in such a situation. Concerning the characteristics of the fixated agents, results showed that human gaze, during navigation, is attracted by dangerous individuals: they were the ones presenting the higher risk of future collision with the participants. Future work is needed to evaluate the influence of other factors such as walking speed or the nature of agents trajectories. This year, we also performed an important experimental campaign including 80 participants to investigate collective behavior (Figure 4.e). When people walk together in the street, they have to coordinate their own motion with the ones of their neighbors. From these local interactions, group motion emerges. The objective of this study was to understand how a collective behavior can emerge from these local interactions between individuals. Especially, the study aimed at identifying what is the neighborhood of a walker in a group from a perceptual point of view (who influences your motion). This work was performed in collaboration with William Warren (Brown University, Providence) and Cécile Appert-Rolland (CNRS, Orsay). Data analysis is still in process but from these results we hope to develop new knowledge on pedestrian behavior. These new results will help us to design new or improve existing crowd simulators based on local interactions. These simulators have important economic and societal roles. For example, they allow to validate the design of public places/building, which aims at hosting dense levels of public in perfectly safe conditions. The study of multiple interactions was also strengthened with the arrival of Laurentius Meerhoff as a post-doctoral student with a regional SAD funding in May 2016. Experiments involving 3 walkers were conducted (Figure 4.f). We investigated how collision is avoided in small groups of people and whether people can successfully interact with the whole environment, or whether under some circumstance agents had to resort to sequential treatment. We proposed a method to detect whether the treatment was sequential or simultaneous and we showed the initial relative position between walkers strongly affects how interaction is engaged with.

Third, we started working on the interaction between a walker and a person on a motorized wheelchair (Figure 4.g). This work was performed in collaboration with the Inria Lagadic team. The main objective was to design a control law that allows the wheelchair to automatically navigate in a crowded place without any collision. This is important for people who have difficulties to drive their wheelchair because of cognitive impairments. However, before reaching this objective, some steps are required to understand how walkers and persons on a wheelchair interact together. To this end, we developed a study where we recorded the trajectory of walkers and a person on a wheelchair in a collision avoidance and reaching scenario. Results will help to model such a control law for natural interactions.

Finally, we continue working on the interaction between a walker and a moving robot. This work was performed in collaboration with Philippe Souères and Christian Vassallo (LAAS, Toulouse). The development of Robotics accelerated these recent years, it is clear that robots and humans will share the same environment in a near future. In this context, understanding local interactions between humans and robots during locomotion tasks is important to steer robots among humans in a safe manner. Our work is a first step in this direction. Our goal is to describe how, during locomotion, humans avoid collision with a moving robot. We just published in Gait and Posture our results on collision avoidance between participants and a non-reactive robot (we wanted to avoid the effect of a complex loop by a robot reacting to participants' motion). Our objective was to determine whether the main characteristics of such interaction preserve the ones previously observed: accurate estimation of collision risk, anticipated and efficient adaptations. We observed that collision avoidance between a human and a robot has similarities with human-human interactions (estimation of collision risk, anticipation) but also leads to major differences [18]. Humans preferentially give way to the robot, even if this choice is not optimal with regard to motion adaptation to avoid the collision. In this new study, we considered the situation where the robot was reactive to the walker's motion (Figure 4.h). First of all, it results that humans have a good understanding of the robot behavior and their reaction are smoother and faster with respect to the case with a non-collaborative robot. Second, humans adapt similarly to human-human study and the crossing order is respected in almost all cases. These results have strong similarities with the ones observed with two humans crossing each other.

7.2.5. *Biomechanics for motion analysis-synthesis*

Participants: Charles Pontonnier, Georges Dumont, Ana Lucia Cruz Ruiz, Antoine Muller, Diane Haering.

In the context of Ana Lucia Cruz Ruiz's PhD, whose goal is to define and evaluate muscle-based controllers for avatar animation, we developed an original control approach to reduce the redundancy of the musculoskeletal system for motion synthesis, based on the muscle synergy theory. For this purpose we ran an experimental campaign of overhead throwing motions. We recorded the muscle activity of 10 muscles of the arm and the motion of the subjects. Thanks to a synergy extraction algorithm, we extracted a reduced set of activation signals corresponding to the so called muscle synergies and used them as an input in a forward dynamics pipeline. Thanks to a two stage optimization method, we adapted the model's muscle parameters and the synergy signals to be as close as possible of the recorded motion. The results are compelling and ask for further developments [5]. We also proposed a classification about muscle-based controllers for animation that has been published in Computer Graphics Forum [6]. Ana Lucia defended her thesis on December 2nd, 2016.

We are also developing an analysis pipeline thanks to the work of Antoine Muller. This pipeline aims at using a modular and multiscale description of the human body to let users be able to analyse human motion. For now, the pipeline is able to assemble different biomechanical models in a convenient descriptive graph, to calibrate these models thanks to experimental data and to compute inverse dynamics to get joint torques from experimental motion capture data. We also investigated the capacity of motion-based methods to calibrate body segment inertial parameters in characterizing the part of the residuals due to the kinematical error into the dynamical one [23].

We also begin to work on the determination of maximal torque envelopes of the elbow thanks to the work of Diane Haering, Inria Post-doctoral fellow at MimeTIC. These results have a great potential of application i) to quantify the articular load during work tasks and ii) to help calibrating muscle parameters for musculoskeletal simulations. Preliminary results have been presented to an international biomechanics conference [21].

Finally, in collaboration with the CERAH (Centre d'étude et d'appareillage des handicapés, institut des invalides, Créteil, France), we proposed an identification-based method for knee prosthesis characteristics. The method is based on a forward dynamics framework enabling a matching between experimental data and model behavior [26].

7.3. Virtual Human Simulation

In addition to this last contribution on biomechanically-inspired character simulation, at the crossroad between motion analysis and simulation, we also explored two main directions for virtual human simulation in 2016. Firstly, with the arrival of Antonio Mucherino in the team, we pushed the idea of extending the idea of interaction meshes (introduced in 2010 by Taku Komura in Edinburgh) to model the constraints intrinsically associated with the motion. This approach requires developing new distance geometry algorithms in order to take time and rigid body constraints into account. Secondly, we continued to push the idea of using perceptual studies to efficiently adapt simulation in order to save computation time for less important details.

Julien Pettré moved to the Lagadic Inria team in March 2016. However we continue collaborating with him on crowd simulation problems, e.g., developing models related to interactions between pedestrians and designing perceptual studies to improve the realism of simulations.

7.3.1. *Recent advances in discretizable distance geometry*

Participants: Antonio Mucherino, Ludovic Hoyet, Franck Multon.

Since September 2016, Antonio Mucherino has a half-time Inria detachment in the MimeTIC team, in order to collaborate on exploring distance geometry-based problems in representing and editing human motion. In collaboration with various French and international partners, he has been working on the different facets of the discretization of the distance geometry. In 2016, he has mainly focused on the two following points. Firstly, since the discretization assumptions require the existence of a vertex ordering on the graph G which is used for representing a problem instance, he presented a new algorithm for the automatic detection of vertex orders that are also able to optimize a given set of objectives [7]. With the aim of making its exploration more efficient, the idea is to reduce in size the search space obtained with the discretization, while keeping in its interior the entire solution set. Secondly, he has started to investigate the possibility to extend the distance geometry (and its discretization) to a wider range of applications, by studying the overlaps between two different geometrical applications, arising in two different domains [3].

More related to the integration with the work in MimeTIC, we are currently exploring applying distance geometry approaches to other applications of interest for virtual human simulations, such as human motion editing and retargeting, and crowd simulations.

7.3.2. Perception of Secondary Motions in Crowd Scenarios

Participants: Ludovic Hoyet, Anne-Hélène Olivier, Richard Kulpa, Julien Pettré.

Creating plausible virtual character animations is of importance in topics researched in MimeTIC, especially for interactive applications where balancing realism and computational load is a requisite. Recently, we investigated how to improve realism of virtual crowd animations by exploring the effects of introducing secondary shoulder motions at the animation level. Typically, a crowd engine pipeline animates numerous moving characters according to a two-step process. First, a crowd simulator generates the characters' global 2D displacement trajectories in the environment, then an animation engine transforms these global trajectories into full body motions. This two-step decomposition is interesting for computational reasons, as crowd simulators raise quadratic complexity issues by nature. For the sake of simplicity, simulation models are often limited to 2D moving circles with 3 degrees of freedom (DoF), i.e., two translations and a rotation. The complete set of internal trajectories (30 to 60 DoF per character) is then considered at the animation step only, where characters are processed independently. This two-step process avoids combining the complexity of crowd simulators with the dimensionality of character kinematic models. However, it also leads to the notion of interactions between characters to be considered only at the simulation level, and to be lost at the animation level. Body animations are therefore not influenced by the presence of neighbours, only global trajectories are. Final animations therefore often lead to residual collisions and/or characters walking as if they were alone, showing no sign to the influence of others.

In this work, we investigated the value of adding secondary motions on the perceived visual quality of crowd animations (i.e., perceived residual collisions and animation naturalness). We focused on adding shoulder motions to characters passing at close distances, and explored this question through two perceptual experiments. To understand the effects of shoulder motions on walking interactions, we first focused on understanding how these secondary motions affect how viewers perceive local interactions between two characters. We found that shoulder motions have strong positive effects on the visual quality of two-character animations, where such animations are perceived to be significantly more natural, and residual collisions become significantly less perceptible. Then we evaluated the benefits of displaying shoulder motions in the situation of crowded scenes, where shoulder motions are diluted into much more visually complex animations, and demonstrated positive effects on the animation naturalness. This increase of visual quality is obtained at a very low computational overhead, which demonstrates the relevance of the direction explored by our work. Our general conclusion is that adding secondary motions in character interactions has a significant impact on the visual quality of crowd animations, with a very light impact on the computational cost of the whole animation pipeline. Our results advance crowd animation techniques by enhancing the simulation of complex interactions between crowd characters with simple secondary motion triggering techniques.

These results were accepted and presented in SIGGRAPH 2016, the premier and most selective computer graphics scientific event, and published in ACM Transaction on Graphics [11].

7.4. Human Motions in VR

To carry-out natural and efficient interactions with a digital world, it is firstly necessary to recognize and evaluate the action of the user. We consequently initiated a collaboration with the Intuidoc IRISA team to adapt methods previously used in 2D gesture recognition to 3D motion. With the increasing use of head mounted display devices (especially cheap devices recently spread in the large public), the problem of avatar simulation and embodiment has become an important challenge. In this context, we initiated collaborative works with Hybrid to better understand embodiment and consequently imagine the future generation of avatars. Concurrently, we continued to explore the use of such technology in various application domains where human performance is a key point, such as ergonomics.

7.4.1. Motion recognition and classification

Participants: Franck Multon, Richard Kulpa, Yacine Boulahia.

Action recognition based on human skeleton structure represents nowadays a prospering research field. This is mainly due to the recent advances in terms of capture technologies and skeleton extraction algorithms. In this context, we observed that 3D skeleton-based actions share several properties with handwritten symbols since they both result from a human performance. We accordingly hypothesize that the action recognition problem can take advantage of trial and error approaches already carried out on handwritten patterns. Therefore, inspired by one of the most efficient and compact handwriting feature-set, we proposed a skeleton descriptor referred to as Handwriting-Inspired Features [20]. First of all, joint trajectories are preprocessed in order to handle the variability among actor's morphologies. Then we extract the HIF3D features from the processed joint locations according to a time partitioning scheme so as to additionally encode the temporal information over the sequence. Finally, we used Support Vector Machine (SVM) for classification. Evaluations conducted on two challenging datasets, namely HDM05 and UTKinect, testify the soundness of our approach as the obtained results outperform the state-of-the-art algorithms that rely on skeleton data.

This work has been carried-out in collaboration with the IRISA Intuidoc team, with Yacine Boulahia who is a co-supervised PhD student with Eric Anquetil.

7.4.2. Avatar Embodiment in Virtual Reality

Participant: Ludovic Hoyet.

With the massive development of virtual reality products investigated by major industrial companies (Google, Facebook, HTC, Sony, etc), there is a new need for understanding what makes users immersed in virtual environments, especially regarding their relation to their virtual representation (i.e., avatar). Amongst others, an important factor is for users to feel incarnated in their avatar, which is called *virtual embodiment*. As more and more technological limitations are now being unlocked, understanding such factors become important to lever new immersive applications, e.g., in education, ergonomics or entertainment.

In collaboration with the EPI Hybrid (Ferran Argelaguet and Anatole Lécuyer), we explore the capacity of avatars to convey such a sense of "virtual embodiment", i.e., the extent to which we accept an avatar to be our representation in the virtual environment. The question of embodiment originates from the famous Rubber Hand Illusion experiment of Botvinick and Cohen (1998). This experiment demonstrated that when participants are presented with a fake rubber hand positioned beside their real hidden hand, and that both hands are synchronously stroked by an experimenter, after some time participants consider their real hand to be positioned at the location of the fake rubber hand. Today, understanding how similar phenomena happen in virtual environments is crucial to provide a maximum immersion for users. For instance, previous work demonstrated that racial biases can be reduced when users are incarnated in virtual characters of a different race, or explored body weight perception by altering the morphology of the avatar. The innovative aspect of our contributions is that we explore this embodiment effect in terms of interactions of the user with the virtual environment.

So far, we explored how people appropriate avatars by evaluating how they accept different representations of their virtual hand in virtual environments. Using various representations ranging from simplistic to highly realistic when interacting in virtual environments [19], we demonstrated that the sense of ownership (i.e., the impression that the virtual hand is actually our own hand) is increased when displaying highly realistic hand representations, but that the sense of agency (i.e., the impression to be able to control the actions of the virtual hand) is stronger for less realistic representations. With the potential of VR to alter and control avatars in different ways, e.g., the user representation, we also explored how structural differences of the hand representation can influence embodiment through controlling a six-digit virtual hand [10]. We found that participants responded positively to the possibility of controlling the virtual hand despite the structural difference, and accepted it as their own to some extent. Overall, results from such experiments further our understanding of the capacity of avatars to elicit a sense of embodiment in the users, and help to design more immersive VR experiences.

7.4.3. VR and Ergonomics

Participants: Charles Pontonnier, Georges Dumont, Pierre Plantard, Franck Multon.

The use of virtual reality tools for ergonomics applications is a very important challenge in order to generalize the use of such devices for the design of workstations.

We proposed a framework for collaborative ergonomic design in virtual environments. The framework consists in defining design modes and metaphors that help the users (engineers, ergonomists, end-users) to find a good trade-off between their own design constraints that can be contradictory at some point. We evaluated the framework and concluded that the active user has to be carefully chosen with regard to the design specifications, since the active user is favouring systematically its own constraints. This work has been published in the Journal on Multimodal User Interfaces [14].

7.5. Digital Storytelling

A transversal research of MimeTIC is digital storytelling as it enables to analyse, capture, model and simulate scenarios involving several humans (real and/or virtual). In this context, it is important to propose annotation tools and languages being able to capture such scenarios and stylistic informations before being able to simulate new ones. Moreover, when living an immersive experience in VR the user may want to have a summarize of his experience, which goes beyond simply replaying the recorded motions. Narration techniques can be positively used to highlight key events and actions, with nonlinear storytelling and intelligent camera placement to convey the desired emotion. The research in this field in MimeTIC contributes to the creation of complex stories on social and human themes. Such approaches are more and more required to create interactive storylines, which massively enhances the possibilities of interactive entertainment, training, computer games and digital applications.

7.5.1. Trip Synopsis: virtual camera control applied to route visualisation

Participant: Marc Christie.

Computerized route planning tools are widely used today by travelers all around the globe, while 3D terrain and urban models are becoming increasingly elaborate and abundant. This makes it feasible to generate a virtual 3D flyby along a planned route. Such a flyby may be useful, either as a preview of the trip, or as an after-the-fact visual summary. However, a naively generated preview is likely to contain many boring portions, while skipping too quickly over areas worthy of attention. We have therefore proposed a general interest-driven framework that automatically computes a flyby along a planned route [9]. This flyby relies on an interest function to derive how close and how slow the camera should focus on the interesting areas, while skipping interest-less regions by using elevated smoothed camera motions. To address the problem, we devised a specific iterative solving process that incrementally approaches the optimal camera trajectory by adjusting position and speed.

7.5.2. Flashbacks in narratives

Participants: Marc Christie, Hui-Yin Wu.

The flashback is a well-known storytelling device used to invoke surprise, suspense, or fill in missing details in a story. Film literature provides a deeper and more complex grounding of flashbacks by explaining their role to stimulate the viewer's memory in order to guide and change viewer comprehension. Yet, in adapting flashback mechanisms to AI storytelling systems, existing approaches have not fully modelled the roles of a flashback event on the viewer's comprehension and memory. To expand the scope of AI generated stories, we propose a formal definition of flashbacks based on the identification of four different impacts on the viewer's beliefs. We then establish a cognitive model that can predict how viewers would perceive a flashback event. We finally design a user evaluation to demonstrate that our model correctly predicts the effects of different flashbacks. This opens great opportunities for creating compelling and temporally complex interactive narratives grounded on cognitive models [29].

7.5.3. Embedded Cinematography Patterns for film Analysis

Participants: Marc Christie, Hui-Yin Wu.

Cinematography carries messages on the plot, emotion, or more general feeling of the film. Yet cinematographic devices are often overlooked in existing approaches to film analysis. To solve this limitation, we present Embedded Constrained Patterns (ECPs), a dedicated query language to search annotated film clips for sequences that fulfill complex stylistic constraints [28]. ECPs are groups of framing and sequencing constraints defined using vocabulary in film textbooks. Using a set algorithm, all occurrences of the ECPs can be found in annotated film sequences. We use a film clip from the Lord of the Rings to demonstrate a range of ECPs that can be detected, and analyse them in relation to story and emotions in the film.

8. Bilateral Contracts and Grants with Industry

8.1. Cifre Faurecia

Participants: Franck Multon [contact], Pierre Plantard.

This contract aims at developing new ergonomics assessments based on inaccurate Kinect measurements in real manufacturing conditions. The main challenges are:

- being able to improve the Microsoft Kinect measurement in order to extract accurate poses from depth images while occlusions may occur,
- developing new inverse dynamics methods based on such inaccurate kinematic data in order to estimate the joint torques required to perform the observed task,
- and proposing a new assessment tool to translate joint torques and poses into potential musculoskeletal disorders risks.

Faurecia has developed its own assessment tool but it requires tedious and subjective tasks for the user, at specific times in the work cycle. By using Kinect information we aim at providing more objective data over the whole cycle not only for specific times. We also wish to make the user focus on the interpretation and understanding of the operator's tasks instead of taking time estimating joint angles in images.

This work is performed in close collaboration with an ergonomist in Faurecia together with the software development service of the company to design the new version of their assessment tool. This tool will be first evaluated on a selection of manufacturing sites and will then be spread worldwide among the 300 Faurecia sites in 33 countries.

This contract enabled us to hire Pierre Plantard as a PhD student to carry-out this work in MimeTIC and M2S Lab. He started in January 2013, finished at the beginning of 2016, and defended his PhD in July 2016. This contract was the opportunity to demonstrate the impact of MimeTIC's work about in-site motion capture on ergonomic assessment, as a decision-support system for ergonomists. The software Kimea is one of the results of this collaboration. It is currently spread in the factories of Faurecia around the world, which demonstrates the maturity of this work for industrial transfer. The method has been published with ergonomic validation in the famous journal Applied Ergonomics (see Highlight section).

9. Partnerships and Cooperations

9.1. National Initiatives

9.1.1. ANR

9.1.1.1. Cineviz

Participants: Marc Christie [contact], Christophe Lino, Hui-Yin Wu.

Cineviz is a 3-year ANR LabCom project (2016-2019). Amount: 300kE. Parnters: SolidAnim, UR1.

The project is a bilateral collaboration with the SolidAnim company. The objective is to jointly progress on the design and implementation of novel tools for the preproduction in the film industry. The project will address the challenges related to (i) proposing expressive framing tools, (ii) integrating the technical aspects of shooting (how to place the cameras, lights, green sets) directly at the design stage), and (iii) novel interaction metaphors for designing and controlling the staging of lights in preproduction, using an example-based approach.

9.1.1.2. Cinecitta

Participants: Marc Christie [contact], Christophe Lino, Hui-Yin Wu.

Cinecitta is a 3.5 year ANR young researcher project lead by Marc Christie (ANR JCJC 2012-2016). Amount: 208kE.

The main objective of Cinecitta was to propose and evaluate a novel workflow which mixes user interaction using motion-tracked cameras and automated computation aspects for interactive virtual cinematography that will better support user creativity. We designed a novel cinematographic workflow that features a dynamic collaboration of a creative human filmmaker with an automated virtual camera planner. The process enhances the quality and utility of the automated planner's suggestions by adapting and reacting to the creative choices made by the filmmaker. This required three advances in the field. First, the ability to generate relevant viewpoint suggestions following classical cinematic conventions. The formalization of these conventions in a computationally efficient and expressive model is a challenging task in order to select and propose the user with a relevant subset of viewpoints among millions of possibilities. Second, the ability to analyze data from real movies in order to formalize some elements of cinematographic style and genre. Third, the integration of motion-tracked cameras in the workflow. Motion-tracked cameras represent a great potential for cinematographic content creation. However given that tracking spaces are of limited size, there is a need to provide novel interaction metaphors to ease the process of content creation with tracked cameras. Finally we gathered feedback on our prototype by involving professionals (during dedicated workshops) and numerous interactions with the Louis Lumière Film School.

9.1.1.3. Entracte

Participants: Charles Pontonnier [contact], Georges Dumont, Franck Multon, Pierre Plantard, Ana Lucia Cruz Ruiz, Antoine Muller, Anthony Sorel, Nicolas Bideau, Richard Kulpa.

The ANR project ENTRACTE is a collaboration between the Gepetto team in LAAS, Toulouse (head of the project) and the Inria/MimeTIC team. The project started in November 2013 and will end in August 2017. The purpose of the ENTRACTE project is to address the action planning problem, crucial for robots as well as for virtual human avatars, in analyzing human motion at a biomechanical level and in defining from this analysis bio-inspired motor control laws and bio-inspired paradigms for action planning. The project is launched since november 2013 and Ana Lucia Cruz Ruiz, who has been recruited as a PhD student since this date, just defended her thesis on muscle-based control based on synergies.

9.1.2. National scientific collaborations

9.1.2.1. Cavaletic

Participant: Franck Multon.

The Cavaletic collaborative project is leaded by University Bretagne Sud and also involves University Rennes2 (CREAD Lab.). It has been funded by the National IFCE (Institut Français du Cheval et de l'Equitation) in order to develop and evaluate technological assistance in horse riding learning, thanks to a user-centered approach. MimeTIC is involved in measuring expert and non-expert horse riders' motions in standardized situations in order to develop metrics to measure riders' performance. It will be used to develop a technological system embedded on users to evaluate their performance and provide them with real-time feedback to correct potential errors.

9.1.2.2. FFT

Participants: Richard Kulpa, Benoit Bideau, Pierre Touzard.

An exclusive contract has been signed between the M2S laboratory and the French Federation of Tennis for three years. The goal is to perform biomechanical analyses of 3D tennis serves on a population of 40 players of the Pôle France. The objective is to determine the link between injuries and biomechanical constraints on joints and muscles depending on the age and gender of the players. At the end, the goal is to evaluate their load training.

9.1.2.3. gDGA

Participants: Antonio Mucherino, Ludovic Hoyet, Franck Multon.

gDGA (generalization of the Distance Geometry and its Applications) is a INS2I/CNRS PEPS project involving local and national partners. Distance geometry can nowadays be seen as a classical problem in operational research, having a wide range of applications. The main aim of this interdisciplinary project is to extend the definition and the range of applicability of distance geometry. In particular, our main interest is on dynamical problems, motivated by a certain number of applications of interest, including interaction motion adaptation, the simulation of crowd behaviours, and the conception of modern recommender systems. The classical application of distance geometry arising in the biological field is also taken into consideration. The necessity of a strong computational power for the considered applications motivates the need of implementing our algorithms in environments capable of exploiting the resources on GPU cards.

9.1.2.4. IRMA

Participants: Ronan Gaugne [contact], Georges Dumont.

The IRMA project is an Imag'In project funded by CNRS which aims at developping innovative methodologies for research in the field of cultural heritage based on the combination of medical imaging technologies and interactive 3D technologies (virtual reality, augmented reality, haptics, additive manufacturing). It relies on close collaborations with the National Institute of Preventive Archaeological Research (Inrap), the Research Center Archaeology, and History Archéosciences (CReAAH UMR 6566) and the company Image ET. The developed tools are intended for cultural heritage professionals such as museums, curators, restorers, and archaeologists. We focus on a large number of archeological artefacts of different nature, and various time periods (Paleolithic, Mesolithic, and Iron Age Medieval) from all over France. We can notably mention the oldest human bones found in Brittany (clavicle Beg Er Vil), a funeral urn from Trebeurden (22), or a Bronze Cauldron from a burial of the Merovingian necropolis "Crassés Saint-Dizier" (51). This project involves a strong collaboration with members of the team Hybrid (Valérie Gouranton, Bruno Arnaldi and Jean-Baptiste Barreau), Théophane Nicolas (Inrap/UMR Trajectoires), Quentin Petit (SED Inria Rennes), and Grégor Marchand (CNRS/UMR CReAAH).

9.1.3. ADT: Immerstar

Participants: Franck Multon, Georges Dumont, Ronan Gaugne.

The ADT-Immerstar is driven by the SED and aims at developing new tools and facilities for the scientific community in order to develop demos and use the two immersive rooms in Rennes: Immersia and Immermove. The engineer (Quentin Petit, SED) has the responsibility of homogenizing the software modules and development facilities in each platform, of installing new upgrades and of developping collaborative applications between the two sites.

9.1.4. PRE

Participants: Franck Multon, Ludovic Hoyet.

The Inria PRE entitled "Smart sensors and novel motion representation breakthrough for human performance analysis" aims at designing a new description for human motion in order to automatically capture, measure and transfer the intrinsic constraints of human motion. Current approaches consist in manually editing the constraints associated with a motion, to use classical skeleton representation with joint angles based on direct or indirect measurements, and then perform inverse kinematics to fulfill these constraints. We aim at designing a new representation to simplify this process pipeline and make it automatic, together with relevant motion sensors that could provide enough information to automatically extract these intrinsic constraints. To this end, this project has been jointly proposed with the Inria CAIRN team, which develops sensors based on joint orientations and distances between sensors. We aim at extending this type of device to measure new types of information that would help to simplify the above mentionned pipeline. Zhiguang Liu started to work as a research fellow on this project since November 2016, working in collaboration with CAIRN. We also involved Hubert Shum from Northumbria University to link this project with our long-term collaboration on this type of problems.

9.2. International Initiatives

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

9.2.1.1. FORMOSA

Title: Fostering Research on Models for Storytelling Applications

International Partner (Institution - Laboratory - Researcher):

NCCU (Taiwan) - Intelligent Media Lab (IML) - Tsai-Yen Li

Start year: 2016

See also: http://www.irisa.fr/mimetic/GENS/mchristi/EA-FORMOSA/

Interactive Storytelling is a new media which allows users to alter the content and outcome of narratives through role-playing and specific actions. With the quality, the availability and reasonable costs of display technologies and 3D interaction devices on one side, and the accessibility of 3D content creation tools on the other, this media is taking a significant share in entertainment (as demonstrated by the success of cinematographic games such as Heavy Rain or Beyond: two souls). These advances push us to re-think the way narratives are traditionally structured, explore new interactive modalities and provide new interactive cinematographic experiences. As a sequel of the first associate team FORMOSA 1, we propose to address new challenges pertained to interactive storytelling such as the use of temporal structures in narratives, interaction modalities and their impact in terms of immersion, and the adaptation of cinematographic real data to 3D environments. To achieve these objectives, the associate team will rely on the complementary skills of its partners and on the co-supervision of students.

9.2.1.2. RE-SIMS

Title: REal data against crowd SImulation algorithMS

International Partner (Institution - Laboratory - Researcher):

University of North Carolina at Chapel Hill (United States) - GAMMA Research Group (GAMMA) - Ming LIN

Start year: 2015

See also: http://www.irisa.fr/mimetic/GENS/jpettre/EASIMS/easims.html

RE-SIMS aims at gathering the best international research teams working on crowd simulation to allow significant progresses on the level of realism achieved by crowd simulators. To this end, RE-SIMS aims at improving methods for capturing crowd motion data that describe real crowd behaviors, as well as by improving data assimilation techniques.

In this renewal, RE-SIMS extends the previous SIMS partnership and follows a multidiciplinary direction.

9.2.2. Informal Inria International Partners

Dr. Edouard Auvinet, Imperial College London, UK (collaboration with Franck Multon, visited the team for a wekk in November)

Dr. Douglas S. Gonçalves, Federal University of Santa Catarina, Florianópolis, Brazil (collaboration with Antonio Mucherino, visited the team in December)

Prof. Carlile Lavor, UNICAMP, Campinas, São Paulo, Brazil (collaboration with Antonio Mucherino)

Dr. Rachel McDonnell, Trinity College Dublin, Ireland (collaboration with Ludovic Hoyet, joint paper submission)

Prof. Carol O'Sullivan, Trinity College Dublin, Ireland (collaboration with Ludovic Hoyet, visited the team for a week in June)

Dr. Huber Shum, Northumbria University, Newcastle, UK (collaboration with Franck Multon and Ludovic Hoyet, with joint papers and supervision, visited the team in November)

9.3. International Research Visitors

9.3.1. Visits of International Scientists

Dr. Edouard Auvinet, Imperial College London, UK (one week in November)

Dr. Douglas S. Gonçalves, Federal University of Santa Catarina, Florianópolis, Brazil (one week in December)

Prof. Carol O'Sullivan, Trinity College Dublin, Ireland (one week in June)

Dr. Hubert Shum, Northumbria University, Newcastle, UK (joint supervision, visit for two days in November)

9.3.1.1. Internships

Yihun Shen, Northumbria University, Newcastle, UK (PhD supervisor: Dr. Hubert Shum), 4-month internship on Rennes Metropole incoming mobility funding (Sept. to Dec. 2016).

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events Organisation

10.1.1.1. General Chair, Scientific Chair

Marc Christie, Program chair of Motion in Games 2015 (but held in 2016)

10.1.1.2. Member of the Organizing Committees

Antonio Mucherio, organiser of Distance Geometry Day 2016, Rennes, France, December 2016

Anne-Hélène Olivier, co-chair of Workshop VHCIE 2016, IEEE VR 2016, Greenville, United-States, October 2016

Marc Christie, steering committee of Motion in Games 2015 and 2016, co-chair of WICED 2016 (Eurographics Workshop on Intelligent Cinematography and Editing)

10.1.2. Scientific Events Selection

10.1.2.1. Member of the Conference Program Committees

Marc Christie

ACM Motion in Games MIG 2016, San Francisco, United-States, October 2016

ACM Symposium on Computer Animation 2016, Zurich, Switzerland, July 2016

Ludovic Hoyet

ACM Motion in Games MIG 2016, San Francisco, United-States, October 2016

ACM Symposium on Applied Perception 2016, Los Angeles, United-States, July 2016

International Conference on Computer Graphics Theory and Applications 2017, Porto, Portugal, February 2017

Richard Kulpa

International Conference on Computer Graphics Theory and Applications 2017, Porto, Portugal, February 2017

Franck Multon

ACM Motion in Games MIG 2016, San Francisco, United-States, October 2016

Computer Animation and Social Agents CASA 2016, Geneva, Switzerland, May 2016

Anne-Hélène Olivier

ACM Motion in Games MIG 2016, San Francisco, United-States, October 2016

10.1.2.2. Reviewer

Marc Christie

Eurographics, Lyon, France, April 2017

ACM CHI, Denver, United-States, May 2017

ACM Siggraph, Anaheim, United-States, July 2016

Eurographics 2017, CHI 2017, Siggraph 2017,

Ludovic Hoyet

ACM Siggraph, Anaheim, United-States, July 2016

ACM SigChi, San Jose, United-States, May 2016

Eurographics, Lyon, France, April 2017

IEEE VR 2017, Los Angeles, United-States, March 2017

Pacific Graphics, Okinawa, Japan, October 2016

Richard Kulpa

IEEE International Conference on Automatic Face and Gesture Recognition, 2017

Franck Multon

IEEE VR 2017, Los Angeles, United-States, March 2017

Anne-Hélène Olivier

IEEE VR 2017, Los Angeles, United-States, March 2017

ACM Siggraph, Anaheim, United-States, July 2016

ACM Siggraph Asia 2016, Macao, China, December 2016

Charles Pontonnier

ACM VRST 2016, Munich, Germany, November 2016

10.1.3. Journal

10.1.3.1. Member of the Editorial Boards

Marc Christie, Associate Editor of the Visual Computer

Armel Crétual, Editorial board of Journal of Electromyography and Kinesiology

Franck Multon, Presence, MIT Press

10.1.3.2. Reviewer - Reviewing Activities

Marc Christie

IEEE Transactions on Visualization and Computer Graphics, The Visual Computer, Computer Graphics Forum

Armel Crétual

Journal of Electromyography and Kinesiology, British Medical Journal open, Clinical Biomechanics, Journal of Orthopaedic Research, IEEE Computer Graphics and Applications

Ludovic Hoyet

IEEE Transactions on Visualization and Computer Graphics, IEEE Computer Graphics & Applications, Computer & Graphics, The Visual Computer Journal, Computer Animation and Virtual Worlds

Richard Kulpa

Journal of Sports Engineering and Technology, Computer

Antonio Mucherino

Mathematical Programming Series A and B, Springer, October 2016.

Franck Multon

Journal Applied Ergonomics, IEEE Transaction on Haptics, IEEE Trans. on Visualization and Computer Graphics, ACM Transactions on Applied Perception, IEEE Computer Graphics and Applications, International Journal of Computer Games Technology

Anne-Hélène Olivier

Journal of Experimental Psychology Human Perception and Performance

Charles Pontonnier

Applied Ergonomics, IEEE Computer Graphics and Applications, MDPI Computers, IEEE Transactions on Visualization and Computer Graphics, ACM Computing surveys

10.1.4. Invited Talks

Ludovic Hoyet. Perception of Biological Human Motion for Plausible Character Animation. Invited speaker at Université Bretagne Sud, team Expression, Vannes, France, June 2016

Antonio Mucherino. Vertex Orders in Distance Geometry. Invited seminar at LIA, University of Avignon, France. Invited by R. Figueiredo. November 2016.

10.1.5. Scientific Expertise

Marc Christie, expert for CIR (credit impot recherche – three cases in 2016)

Georges Dumont, expert for Dutch Research Council VENI (Netherlands) project

Franck Multon, HCERES expert for UMR 7287 AMU/CNRS "Institut des Sciences du Mouvements" in Marseille en 2016-2017, ANR expert, member of the ANR CPDS 4 "Santé Bien-être" through the Allistene national Alliance to design the next ANR call for projects, expert for international NSRC (Canada), CRSNG (Canada) and Dutch Research Council VENI (Netherlands) projects

Charles Pontonnier, Member of the Normalization Work Group "Impact of exoskeletons on work conditions", groupe AFNOR.

10.1.6. Research Administration

Georges Dumont is president of the elected group at scientific council of École Normale Supérieure de Rennes, member of the scientific council of École Normale Supérieure de Rennes

Georges Dumont is scientific head of Immerstar platforms (Immersia + Immermove) jointly for Inria and Irisa Partners

Ludovic Hoyet participated in the local MESR PhD Grant Auditions 2016

Richard Kulpa is member of the University Rennes2 Research steering committee "commission recherche", and Academic Council "CAC"

Fabrice Lamarche is a member of the national council of universities (Conseil National des Universités - CNU) since november 2015

Antonio Mucherinon is the person in charge of International Relationships at ISTIC, and member of the "Commission Affaires Internationales" at Université Rennes 1

Franck Multon, member of the "Conseil Académique" and "Commission Recherche" of University Rennes2, head of Immermove platform in the Immerstar group, member of the D6 "Media and Interactions" department in IRISA, head of MimeTIC team

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

10.2.1.1. Master

Marc Christie

"Multimedia Mobile", Master 2, leader of the module, 32h, Computer Science, University of Rennes 1, France

"Projet Industriel Transverse", Master 2, 32h, leader of the module, Computer Science, University of Rennes 1, France

"Outils pour la Conception d'IHM", Master 2, 32h, leader of the module, Computer Science, University of Rennes 1, France

Armel Crétual

"Méthodologie", leader of the module, 20h, Master 1 M2S, University Rennes2, France

"Biostatstiques", leader of the module, 30h, Master 2 M2S, University Rennes2, France

Georges Dumont

Responsible of the second year of the master Mechatronics, Rennes 1 University and École Normale Supérieure de Rennes, France

"Mechanical simulation in Virtual Reality", 36h, Master Mechatronics, Rennes 1 University and École Normale Supérieure de Rennes, France

"Mechanics of deformable systems", 40h, Master FE, École Normale Supérieure de Rennes, France

"Oral preparation to agregation competitive exam", 20h, Master FE, École Normale Supérieure de Rennes, France

"Vibrations in Mechanics", 10h, Master FE, École Normale Supérieure de Rennes, France

"Multibody Dynamics", 9h, Master FE, École Normale Supérieure de Rennes, France

"Finite Element method", 12h, Master FE, École Normale Supérieure de Rennes, France

Ronan Gaugne

"VR projects", 40h, M1/M2, INSA Rennes, France

Ludovic Hoyet

"Images et Mouvement - IMO", Part "Mouvement Humain", 10h, Master 2 research in computer sciences, University Rennes1, France

Richard Kulpa

"Boucle analyse-modélisation-simulation du mouvement", 27h, leader of the module, Master 2 M2S, Université Rennes 2, France

"Méthodes numériques d'analyse du geste", 27h, leader of the module, Master 2 M2S, Université Rennes 2, France

"Cinématique inverse", 3h, leader of the module, Master 2 M2S, Université Rennes 2, France

Fabrice Lamarche

"Compilation pour l'image numérique", 29h, Master 1, ESIR, University of Rennes 1, France

"Synthèse d'images", 12h, Master 1, ESIR, University of Rennes 1, France

"Synthèse d'images avancée", 28h, Master 1, ESIR, University of Rennes 1, France

"Modélisation Animation Rendu", 36h, Master 2, ISTIC, University of Rennes 1, France

"Jeux vidéo", 26h, Master 2, ESIR, University of Rennes 1, France

Antonio Mucherino

"Parallel Computing" (in English), 30h, M1 Informatique, Université Rennes 1, France

"Introduction aux Systèmes et Réseaux", 42h, M1 BIG (SVE), Université Rennes 1, France

"Programmation Orienté à Objets", 40h, M1 BIG (SVE), Univ. Université Rennes 1, France

"Algorithmes pour les Séquences et les Structures", 12h, M2 BIG (SVE), Université Rennes 1, France

Franck Multon

"Images et Mouvement - IMO", leader of the module, 20h, Master 2 research in computer sciences, University Rennes1, France

"Santé et Performance au Travail : étude de cas", leader of the module, 30h, Master 1 M2S, University Rennes2, France

"Analyse Biomécanique de la Performance Motrice", leader of the module, 30h, Master 1 M2S, University Rennes2, France

"Modélisation et Simulation du Mouvement", leader of the module, 30h, Master 2 M2S, University Rennes2, France

"Connaissances neurophysiologiques et biomécaniques", Mastere "Excellence Operationnelle", INSA Rennes, France

Anne-Hélène Olivier

"Biostatstiques", 18h, Master 2 M2S, University Rennes2, France

"Biostatstiques", 16h, Master 1 2SEP, École Normale Supérieure de Rennes, France

"Contrôle moteur : loi de contrôle de la locomotion", 3h30, Master 1 M2S, Université Rennes 2, France

"Contrôle moteur : Boucle perceptivo-motrice", 3h30, Master 1 M2S, Université Rennes 2, France

"Analyse Biomécanique de la Performance Motrice", 15h, Master 1 M2S, University Rennes2, France

Charles Pontonnier

"Numerical methods", leader of the module, Mechanics, École Spéciale Militaire de Saint-Cyr Coëtquidan, France

"Numerical simulation of mechanical systems", leader of the module, Mechanics, École Spéciale Militaire de Saint-Cyr Coëtquidan, France

"Analytical Mechanics" , Mechanics, École Spéciale Militaire de Saint-Cyr Coëtquidan, France

"Design and control of legged robots", leader of the module, Electronics, École Spéciale Militaire de Saint-Cyr Coëtquidan, France

"Design, simulation and control of mechanical systems", leader of the module, Lecturers training in mechatronics, École Normale Supérieure de Rennes, France

"Musculoskeletal modeling and ergonomics", 30h, Master 1 M2S, University Rennes2, France

10.2.1.2. License level

Marc Christie

"Programmation Impérative 1", leader of the module, University of Rennes 1, France

Armel Crétual

"Analyse cinématique du mouvement", 100h, Licence 1, University Rennes 2, France

Richard Kulpa

"Biomécanique (dynamique en translation et rotation)", 48h, Licence 2, Université Rennes 2, France

"Méthodes numériques d'analyse du geste", 48h, Licence 3, Université Rennes 2, France

"Statistiques et informatique", 15h, Licence 3, Université Rennes 2, France

Ronan Gaugne

"VR projects", 10h, L3, INSA Rennes, France

Fabrice Lamarche

"Initiation à l'algorithmique et à la programmation", 56h, License 3, ESIR, University of Rennes 1, France

"Programmation en C++", 46h, License 3, ESIR, University of Rennes 1, France

"IMA", 24h, License 3, ENS Rennes, ISTIC, University of Rennes 1, France

Antonio Mucherino

"Programmation Impérative 1", 80h, L1, Université Rennes 1, France

Franck Multon

"Ergonomie du poste de travail", Licence STAPS L2 & L3, University Rennes2, France

Anne-Hélène Olivier

"Analyse cinématique du mouvement", 100h , Licence 1, University Rennes 2, France

"Anatomie fonctionnelle", 8h , Licence 2, University Rennes 2, France

"Effort et efficience", 12h , Licence 2, University Rennes 2, France

"Locomotion et handicap", 12h , Licence 3, University Rennes 2, France

"Biomécanique du viellissement", 12h , Licence 3, University Rennes 2, France

Charles Pontonnier

"Numerical control", leader of the module, Electronics, École Inter-Armes de Saint-Cyr Coëtquidan, France

10.2.2. Supervision

10.2.2.1. PhD (defended)

Julien Bruneau, Studying and modeling complex interactions for crowd simulation, University Rennes 1, defended in Nov. 2016, Julien Pettré (Lagadic) & Anne-Hélène Olivier

Ana Lucia Cruz Ruiz, Low-Dimensional Control Representations for Muscle-Based Characters: Application to Overhead Throwing, ENS Rennes, defended in Dec. 2016, Georges Dumont & Charles Pontonnier

Pierre Plantard, Estimation des efforts musculaires à partir de données in situ pour l'évaluation ergonomique d'un poste de travail, defended in July 2016, Franck Multon & Anne-Sophie LePierres

Cunka Sanokho, Data-driven Virtual Cinematography, University Rennes1, defended in Feb. 2016, Marc Christie

Hui-yin Wu, Cinematic discourse for Interactive 3D Storytelling, University Rennes1, defended in Oct. 2016, Marc Christie

10.2.2.2. PhD (in progress)

Yacine Bouhalia, Approche transversale pour l'analyse et la reconnaissance de gestes 2D et 3D, INSA of Rennes, September 2015, Richard Kulpa & Franck Multon & Eric Anquetil

Charles Faure, Stratégies coopératives et compétitives dans des tâches d'interactions physiques multiples, Université Rennes 2, September 2016, Benoit Bideau & Richard Kulpa

Florence Gaillard, Evaluation en situation écologique des capacités fonctionnelles des membres supérieurs d'enfants hémiplégiques, University Rennes 2, December 2015, Armel Crétual & Isabelle Bonan

Karim Jamal, Les effets des stimulations sensorielles par vibration sur les perturbations posturales secondaires à des troubles de la cognition spatiale après un Accident vasculaire Cérébrale, University Rennes 2, September 2016, Isabelle Bonan & Armel Crétual

Sean D. Lynch, Perception visuelle du mouvement humain dans les interactions lors de tâches locomotrices, M2S - University Rennes 2, September 2015, Anne-Hélène Olivier & Richard Kulpa

Marion Morel, Suivi et étude des interactions pour l'analyse des tactiques durant un match de basket-ball, UPMC - University Rennes 2, September 2014, Catherine Achard & Séverine Dubuisson & Richard Kulpa

Antoine Muller, Design of a modular and multiscale musculoskeletal library as a support to motion analysis-synthesis, Ecole normale supérieure, September 2014, Georges Dumont & Charles Pontonnier

Camille Pouliquen, Optimization of musculoskeletal models during the sporting gesture: Cycling application, University Rennes 2, September 2013, Paul Delamarche & Nicolas Bideau

Pierre Touzard, Suivi longitudinal du service de jeunes joueurs de tennis élite : identification biomécanique des facteurs de performance et de risque de blessures, University Rennes 2, September 2014, Benoit Bideau & Richard Kulpa & Caroline Martin

10.2.3. Juries

PhD: Deep Seth, Contribution to the evaluation of muscle fatigue model and recovery model, Centrale Nantes, Rapporteur Franck Multon

PhD: Pamela Carreno Medrano, Analysis and Synthesis of Expressive Theatrical Movements, Université Bretagne Sud, Rapporteur Franck Multon

10.3. Popularization

Franck Multon managed a 4 days Booth to present ergonomic work in MIDEST in Paris, with a plenary session conference about penibility at work, MIDEST December 2016

Franck Multon and Richard Kulpa managed a 2 days Inria Booth to present VR and Sports in Futur en Seine in Paris, June 2016

Ludovic Hoyet, Journées Scientifiques Inria 2016, exposé court en 180s, “Perceptual Effect of Shoulder Motions on Crowd Animations”.

Ludovic Hoyet, Press conference, « Six-Finger Illusion : pour mieux comprendre la perception de notre corps et de notre avatar ». With Ferran Argelaguet and Anatole Lécuyer, Inria, Paris, 12 May 2016, https://www.inria.fr/centre/rennes/actualites/six-finger-illusion

11. Bibliography

Publications of the year

Doctoral Dissertations and Habilitation Theses

[1] D. WOLINSKI.*Microscopic crowd simulation : evaluation and development of algorithms*, Université Rennes 1, January 2016, https://tel.archives-ouvertes.fr/tel-01420105.

Articles in International Peer-Reviewed Journal

[2] E. AUVINET, F. MULTON, V. MANNING, J. MEUNIER, J. P. COBB.*Validity and sensitivity of the Longitudinal Asymmetry Index to detect gait asymmetry using Microsoft Kinect data*, in "Gait and Posture", 2017, vol. 51, 17 [*DOI :* 10.1016/J.GAITPOST.2016.08.022], https://hal.inria.fr/hal-01359845.

[3] S. J. BILLINGE, P. M. DUXBURY, D. S. GONÇALVES, C. LAVOR, A. MUCHERINO.*Assigned and Unassigned Distance Geometry: Applications to Biological Molecules and Nanostructures*, in "Quarterly Journal of Operations Research", December 2016, vol. 14, n° 4, p. 337–376, https://hal.inria.fr/hal-01401745.

[4] G. BRUDER, F. ARGELAGUET, A.-H. OLIVIER, A. LÉCUYER.*CAVE Size Matters: Effects of Screen Distance and Parallax on Distance Estimation in Large Immersive Display Setups*, in "Presence: Teleoperators and Virtual Environments", 2016, vol. 25, n° 1, p. 1 - 16 [*DOI :* 10.1162/PRES_A_00241], https://hal.inria.fr/hal-01388499.

[5] A. L. CRUZ RUIZ, C. PONTONNIER, J. LEVY, G. DUMONT.*A Synergy-Based Control Solution for Overactuated Characters: Application to Throwing*, in "Computer Animation and Virtual Worlds", 2016, https://hal.inria.fr/hal-01377058.

[6] A. L. CRUZ RUIZ, C. PONTONNIER, N. PRONOST, G. DUMONT.*Muscle-Based Control For Character Animation*, in "Computer Graphics Forum", May 2016, https://hal.inria.fr/hal-01317881.

[7] D. S. GONÇALVES, A. MUCHERINO.*Optimal Partial Discretization Orders for Discretizable Distance Geometry*, in "International Transactions in Operational Research", September 2016, vol. 23, n° 5, p. 947–967 [*DOI :* 10.1111/ITOR.12249], https://hal.inria.fr/hal-01402366.

[8] D. S. GONÇALVES, A. MUCHERINO, C. LAVOR, L. LIBERTI.*Recent Advances on the Interval Distance Geometry*, in "Journal of Global Optimization", January 2017, https://hal.inria.fr/hal-01429540.

[9] H. HUANG, D. LISCHINSKI, Z. HAO, M. GONG, M. CHRISTIE, D. COHEN-OR.*Trip Synopsis: 60km in 60sec*, in "Computer Graphics Forum", 2016, vol. 35, p. 107 - 116 [*DOI :* 10.1111/CGF.13008], https://hal.inria.fr/hal-01413388.

[10] L. HOYET, F. ARGELAGUET, C. NICOLE, A. LÉCUYER.*"Wow! I Have Six Fingers!": Would You Accept Structural Changes of Your Hand in VR?*, in "Frontiers in Robotics and AI", March 2016, vol. 3, n° 27 [*DOI :* 10.3389/FROBT.2016.00027], https://hal.inria.fr/hal-01334359.

[11] L. HOYET, A.-H. OLIVIER, R. KULPA, J. PETTRÉ.*Perceptual Effect of Shoulder Motions on Crowd Animations*, in "ACM Transactions on Graphics", July 2016 [*DOI :* 10.1145/2897824.2925931], https://hal.inria.fr/hal-01357713.

[12] C. MARTIN, B. BIDEAU, P. DELAMARCHE, R. KULPA.*Influence of a Prolonged Tennis Match Play on Serve Biomechanics*, in "PLoS ONE", 2016, vol. 11, n° 8, e0159979 [*DOI :* 10.1371/JOURNAL.PONE.0159979], https://hal-univ-rennes1.archives-ouvertes.fr/hal-01397596.

[13] C. MARTIN, R. KULPA, F. EZANNO, P. DELAMARCHE, B. BIDEAU.*Influence of Playing a Prolonged Tennis Match on Shoulder Internal Range of Motion*, in "American Journal of Sports Medicine", 2016, vol. 44, n° 8, p. 2147–2151 [*DOI :* 10.1177/0363546516645542], https://hal-univ-rennes1.archives-ouvertes.fr/hal-01397597.

[14] H. NGUYEN, C. PONTONNIER, S. HILT, T. DUVAL, G. DUMONT.*VR-based Operating Modes and Metaphors for Collaborative Ergonomic Design of Industrial Workstations*, in "Journal on Multimodal User Interfaces", 2016, https://hal.inria.fr/hal-01381243.

[15] P. PLANTARD, H. P. H. SHUM, A.-S. LE PIERRES, F. MULTON.*Validation of an ergonomic assessment method using Kinect data in real workplace conditions*, in "Applied Ergonomics", 2016 [*DOI :* 10.1016/J.APERGO.2016.10.015], https://hal.inria.fr/hal-01393066.

[16] P. PLANTARD, H. SHUM, F. MULTON.*Filtered Pose Graph for Efficient Kinect Pose Reconstruction*, in "Multimedia Tools and Applications", 2016, 24 [*DOI :* 10.1007/S11042-016-3546-4], https://hal.inria.fr/hal-01332711.

[17] M. ROPARS, A. CRÉTUAL, R. KAILA, I. BONAN, H. ANTHONY, H. THOMAZEAU.*Diagnosis and treatment of anteroinferior capsular redundancy associated with anterior shoulder instability using an open Latarjet procedure and capsulorrhaphy*, in "Knee Surgery, Sports Traumatology, Arthroscopy", 2016, vol. 24, n° 12, p. 3756–3764 [*DOI :* 10.1007/S00167-015-3621-9], https://hal.inria.fr/hal-01239752.

[18] C. VASSALLO, A.-H. OLIVIER, P. SOUÈRES, A. CRÉTUAL, O. STASSE, J. PETTRÉ.*How do walkers avoid a mobile robot crossing their way?*, in "Gait and Posture", January 2017, vol. 51, p. 97-103, https://hal.archives-ouvertes.fr/hal-01371202.

International Conferences with Proceedings

[19] F. ARGELAGUET, L. HOYET, M. TRICO, A. LÉCUYER.*The role of interaction in virtual embodiment: Effects of the virtual hand representation*, in "IEEE Virtual Reality", Greenville, United States, March 2016, p. 3-10 [*DOI :* 10.1109/VR.2016.7504682], https://hal.inria.fr/hal-01346229.

[20] S. Y. BOULAHIA, E. ANQUETIL, R. KULPA, F. MULTON.*HIF3D: Handwriting-Inspired Features for 3D Skeleton-Based Action Recognition*, in "23rd International Conference on Pattern Recognition (ICPR 2016)", Cancun, Mexico, IEEE (editor), December 2016, https://hal.archives-ouvertes.fr/hal-01376113.

[21] D. HAERING, C. PONTONNIER, N. BIDEAU, G. NICOLAS, G. DUMONT.*Preliminary report on the effects of patient positioning and testing modes on elbow isokinetic measurement*, in "22nd Congress of the European Society of Biomechanics", Lyon, France, 22nd Congress of the European Society of Biomechanics, July 2016, https://hal.inria.fr/hal-01317876.

[22] M. MOREL, R. KULPA, A. SOREL, C. ACHARD, S. DUBUISSON.*Automatic and Generic Evaluation of Spatial and Temporal Errors in Sport Motions*, in "11th International Conference on Computer Vision Theory and Applications (VISAPP 2016)", Rome, Italy, February 2016, p. 542-551 [*DOI :* 10.5220/0005778505420551], https://hal.archives-ouvertes.fr/hal-01373822.

[23] A. MULLER, C. PONTONNIER, G. DUMONT.*Uncertainty spreading from kinematics to dynamics in multibody human models*, in "22nd Congress of the European Society of Biomechanics", Lyon, France, 22nd Congress of the European Society of Biomechanics, July 2016, https://hal.inria.fr/hal-01317865.

[24] C. PARK, J. S. PARK, S. TONNEAU, N. MANSARD, F. MULTON, J. PETTRÉ, D. MANOCHA.*Dynamically balanced and plausible trajectory planning for human-like characters*, in "20th ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games", Redmond, United States, ACM, February 2016, p. 39-48 [*DOI :* 10.1145/2856400.2856405], https://hal.inria.fr/hal-01290368.

[25] P. PLANTARD, H. SHUM, F. MULTON.*Ergonomics Measurements using Kinect with a Pose Correction Framework*, in "Digital Human Modeling", Montreal, Canada, June 2016, 8, https://hal.inria.fr/hal-01332716.

[26] C. PONTONNIER, C. VILLA, J. BASCOU.*Identifying knee prosthesis characteristics during swing phase through optimization*, in "22nd Congress of the European Society of Biomechanics", Lyon, France, 22nd Congress of the European Society of Biomechanics, July 2016, https://hal.inria.fr/hal-01317869.

[27] E. SAFA, J.-B. BARREAU, R. GAUGNE, W. DUCHEMIN, J.-D. TALMA, B. ARNALDI, G. DUMONT, V. GOURANTON.*Digital and handcrafting processes applied to sound-studies of archaeological bone flutes*, in "International Conference on Culturage Heritage, EuroMed", Nicosia, Cyprus, 2016, vol. 1, n^{o} 10058, p. 184-195, https://hal.archives-ouvertes.fr/hal-01391755.

[28] H.-Y. WU, M. CHRISTIE.*Analysing Cinematography with Embedded Constrained Patterns*, in "WICED - Eurographics Workshop on Intelligent Cinematography and Editing", Lisbon, Portugal, May 2016, https://hal.inria.fr/hal-01413407.

[29] H.-Y. WU, M. YOUNG, M. CHRISTIE.*A Cognitive-Based Model of Flashbacks for Computational Narratives*, in "Proceedings of Artificial Intelligence and Interactive Digital Entertainment 2016", San Francisco, United States, October 2016, https://hal.inria.fr/hal-01413401.

Conferences without Proceedings

[30] S. TONNEAU, A.-A. RAMI, J. PETTRÉ, T. KOMURA, N. MANSARD.*Character contact re-positioning under large environment deformation*, in "Eurographics", Lisbonne, Portugal, eurographics, May 2016, https://hal.archives-ouvertes.fr/hal-01207688.

Scientific Books (or Scientific Book chapters)

[31] T. E. ASSOCIATION (editor). *Eurographics Workshop on Intelligent Cinematography and Editing*, Eurographics Workshop on Intelligent Cinematography and Editing, The Eurographics Association, Lisbonne, Portugal, May 2016, https://hal.inria.fr/hal-01427243.

Other Publications

[32] S. TONNEAU, A. DEL PRETE, J. PETTRÉ, C. PARK, D. MANOCHA, N. MANSARD.*An efficient acyclic contact planner for multiped robots*, February 2016, working paper or preprint [*DOI :* 10.1177/TOBEASSIGNED], https://hal.archives-ouvertes.fr/hal-01267345.

Project-Team MYRIADS

Design and Implementation of Autonomous Distributed Systems

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

IN PARTNERSHIP WITH:
CNRS
Institut national des sciences appliquées de Rennes
Université Rennes 1

RESEARCH CENTER
Rennes - Bretagne-Atlantique

THEME
Distributed Systems and middleware

Table of contents

Project-Team MYRIADS

Creation of the Team: 2010 January 01, updated into Project-Team: 2012 January 01

Keywords:

Computer Science and Digital Science:

1.1.2. - Hardware accelerators (GPGPU, FPGA, etc.)
1.1.4. - High performance computing
1.1.5. - Exascale
1.1.6. - Cloud
1.1.7. - Peer to peer
1.1.9. - Fault tolerant systems
1.1.13. - Virtualization
1.2. - Networks
1.2.4. - QoS, performance evaluation
1.2.5. - Internet of things
1.3. - Distributed Systems
1.6. - Green Computing
2.1.7. - Distributed programming
2.2.3. - Run-time systems
2.3.2. - Cyber-physical systems
2.4.2. - Model-checking
2.6. - Infrastructure software
2.6.1. - Operating systems
2.6.2. - Middleware
2.6.3. - Virtual machines
3.1.2. - Data management, quering and storage
3.1.3. - Distributed data
4.7. - Access control
4.9. - Security supervision
4.9.1. - Intrusion detection
4.9.3. - Reaction to attacks
5.6. - Virtual reality, augmented reality
6.1.3. - Discrete Modeling (multi-agent, people centered)
6.2.6. - Optimization
6.2.7. - High performance computing
7.1. - Parallel and distributed algorithms
7.3. - Optimization

Other Research Topics and Application Domains:

2.3. - Epidemiology
3.1. - Sustainable development
3.2. - Climate and meteorology
4.3. - Renewable energy production

4.4. - Energy delivery
4.4.1. - Smart grids
4.5. - Energy consumption
4.5.1. - Green computing
5.1. - Factory of the future
5.8. - Learning and training
6.1. - Software industry
6.1.1. - Software engineering
6.3. - Network functions
6.3.3. - Network Management
6.4. - Internet of things
6.5. - Information systems
6.6. - Embedded systems
8.1. - Smart building/home
8.2. - Connected city
8.5. - Smart society
9.1. - Education
9.1.1. - E-learning, MOOC
9.1.2. - Serious games
9.4.1. - Computer science
9.6. - Reproducibility
9.7. - Knowledge dissemination
9.7.1. - Open access
9.7.2. - Open data
9.8. - Privacy

1. Members

Research Scientists

Christine Morin [Team leader, Inria, Senior Researcher, HDR]
Deborah Agarwal [Inria, Senior Researcher, Inria International Chair]
Yvon Jégou [Inria, Researcher, until Sep. 2016]
Anne-Cécile Orgerie [CNRS, Researcher]

Faculty Members

Jean-Pierre Banâtre [Univ. Rennes I, Professor Emeritus]
Nikos Parlavantzas [INSA Rennes, Associate Professor]
Jean-Louis Pazat [INSA Rennes, Professor, HDR]
Guillaume Pierre [Univ. Rennes I, Professor]
Martin Quinson [ENS Rennes, Professor, HDR]
Cédric Tedeschi [Univ. Rennes I, Associate Professor]

Technical Staff

Anirvan Basu [Inria]
Julien Lefeuvre [Inria, granted by FP7 ECO2Clouds project, until Jul. 2016]
Manh Linh Pham [Inria, from Apr. 2016]
Sébastien Philippot [Inria]

Arnab Sinha [Inria, until Nov. 2016]
Alexandre Van Kempen [Inria]
David Margery [Inria, Research engineer]
Pascal Morillon [Univ. Rennes I]
Matthieu Simonin [Inria, Research engineer]

PhD Students

Arif Ahmed [Univ. Rennes I, from Oct. 2016]
Ismael Cuadrado Cordero [EMN, until Sep. 2016]
Anna Giannakou [Inria]
Baptiste Goupille-Lescar [Thales, granted by CIFRE]
David Guyon [Univ. Rennes I]
Sabbir Hasan Rochi [INSA Rennes]
Ancuta Iordache [Univ. Rennes I, until Jan. 2016]
Yunbo Li [EMN]
Edouard Outin [Inst. de Recherche Technologique B-COM, until Nov. 2016]
The Anh Pham [Inria, from Nov. 2016]
Genc Tato [Inria]
Amir Teshome Wonjiga [Inria]

Post-Doctoral Fellow

Benjamin Camus [Inria, from Oct. 2016]

Visiting Scientists

Carlos Ruiz Diaz [University of Guadalajara, Mexico, until Feb. 2016]
Palakiyem Wallah [University of Kara, Togo]

Administrative Assistant

Maryse Fouché [Inria]

Others

Simon Bihel [Inria, Intern, from May 2016 until Jul. 2016]
Jean Leon Cusinato [Inria, Intern, from Mar. 2016 until Aug. 2016]
Arnaud Fabre [Univ. Rennes I, Intern, from Feb. 2016 until Jul. 2016]
Yann Jegu [Inria, Intern, from May 2016 until Jul. 2016]
Fergal Martin-Tricot [Inria, Intern, from Jun. 2016 until Aug. 2016]
Louis Rilling [DGA, Researcher-Engineer, Associate Member]
Debaditya Roy [Univ. Rennes I, Intern, from May 2016 until Aug. 2016]
Kartik Sathyanarayanan [Inria, Intern, from May 2016 until Jul. 2016]
Simon Trebern [INSA Rennes, Intern, from Jun. 2016 until Jul. 2016]
Benjamin Trubert [Univ. Rennes I, Intern, from May 2016 until Aug. 2016]

2. Overall Objectives

2.1. General Objectives

MYRIADS is a joint team with INRIA, CNRS, UNIVERSITY RENNES 1, INSA RENNES and ENS RENNES. It is part of IRISA (D1 department on large scale systems) and INRIA RENNES – BRETAGNE ATLANTIQUE.

The objective of MYRIADS is to design and implement systems for autonomous service and resource management in interconnected and distributed clouds. The team tackles the challenges of dependable application execution and efficient resource management in highly distributed clouds.

2.2. Context

The MYRIADS team research activities are conducted in the context of the future of Internet.

Internet of Services. Myriads of applications are provided to more than one billion users [0] all over the world. Over time, these applications are becoming more and more sophisticated, a given application being a composition of services likely to be executed on various sites located in different geographical locations. The Internet of Services is spreading all domains: home, administration, business, industry and science. Everyone is involved in the Internet of Services: citizens, enterprises, scientists are application, service and resource consumers and/or providers over the Internet.

Outsourcing. Software is provided as a service over the Internet. Myriads of applications are available on-line to billions of users as, for instance, *GoogleApps* (Gmail). After decades in which companies used to host their entire IT infrastructures in-house, a major shift is occurring where these infrastructures are outsourced to external operators such as Data Centers and Computing Clouds. In the Internet of Services, not only software but also infrastructure are delivered as a service. Clouds turned computing and storage into a utility. Just like water or electricity, they are available in virtually infinite amounts and their consumption can be adapted within seconds like opening or closing a water tap. The main transition, however, is the change in business models. Companies or scientists do not need to buy and operate their own data centers anymore. Instead, the compute and storage resources are offered by companies on a "pay-as-you-go" basis. There is no more need for large hardware investments before starting a business. Even more, the new model allows users to adapt their resources within minutes, e.g., scale up to handle peak loads or rent large numbers of computers for a short experiment. The risk of wasting money by either under-utilization or undersized data centers is shifted from the user to the provider.

Sharing and Cooperation. Sharing information and cooperating over the Internet are also important user needs both in the private and the professional spheres. This is exemplified by various services that have been developed in the last decade. Peer-to-peer networks are extensively used by citizens in order to share musics and movies. A service like *Flickr* allowing individuals to share pictures is also very popular. Social networks such as *FaceBook* or *Linkedln* link millions of users who share various kinds of information within communities. Virtual organizations tightly connected to Grids allow scientists to share computing resources aggregated from different institutions (universities, computing centers...). The EGEE European Grid is an example of production Grid shared by thousands of scientists all over Europe.

2.3. Challenges

The term cloud was coined 10 years ago. Today cloud computing is widely adopted for a wide range of usage: information systems outsourcing, web service hosting, scientific computing, data analytics, back-end of mobile and IoT applications. There is a wide variety of cloud service providers (IaaS, PaaS, SaaS) resulting in difficulties for customers to select the services fitting their needs. Production clouds are powered by huge data centers that customers reach through the Internet. This current model raises a number of issues. Cloud computing generates a lot of traffic resulting in ISP providers needing to increase the network capacity. An increasing amount of always larger data centers consumes a lot of energy. Cloud customers experience poor quality of experience for highly interactive mobile applications as their requests are dealt with in data centers that are several hops away. The centralization of data in clouds also raises (i) security issues as clouds are a target of choice for attackers and (ii) privacy issues with data aggregation. Recently new cloud architectures have been proposed to overcome the scalability, latency, and energy issues of traditional centralized data centers. Various flavors of distributed cloud computing are emerging depending on the resources exploited: resources in the core network (distributed cloud), resources at the edge of the network (edge clouds) and even resources in the people swarms of devices (fog computing) enabling scalable cloud computing. These distributed clouds raise new challenges for resource and application management.

[0] According to World Stats, there are 3.67 billion Internet users i.e. more than half of the total world population in June 2016 http://www.internetworldstats.com/stats.htm.

The ultimate goal of Myriads team is making highly distributed clouds sustainable. By sustainability we mean green, efficient and secure clouds. We plan to study highly distributed clouds including edge clouds and fog computing. In this context, we will investigate novel techniques for greening clouds including the optimization of energy consumption in distributed clouds in the context of smart grids. As more and more critical information system are outsourced in the cloud and personal data captured by sensors embedded in smart objects and smartphones are stored in the cloud, we will investigate security and privacy issues in two directions: cloud security monitoring and personal data protection in cloud-based IoT applications.

System research requires experimental validation based on simulation and/or prototyping. Reproducible experimentation is essential. We will contribute to the design and implementation of simulators well suited to the study of distributed clouds (architecture, energy consumption) and of large scale experimentation platforms for distributed systems enabling reproducible experiments.

3. Research Program

3.1. Introduction

In this section, we present our research challenges along four work directions: resource and application management in distributed cloud architectures for scaling clouds in Section 3.2, energy management strategies for greening clouds in Section 3.3, security and data protection aspects for securing cloud-based information systems and applications in Section 3.4, and methods for experimenting with clouds in Section 3.5.

3.2. Scaling clouds

3.2.1. Resource management in hierarchical clouds

The next generation of utility computing appears to be an evolution from highly centralized clouds towards more decentralized platforms. Today, cloud computing platforms mostly rely on large data centers servicing a multitude of clients from the edge of the Internet. Servicing cloud clients in this manner suggests that locality patterns are ignored: wherever the client issues his/her request from, the request will have to go through the backbone of the Internet provider to the other side of the network where the data center relies. Besides this extra network traffic and this latency overhead that could be avoided, other common centralization drawbacks in this context stand in limitations in terms of security/legal issues and resilience.

At the same time, it appears that network backbones are over-provisioned for most of their usage. This advocates for placing computing resources directly within the backbone network. The general challenge of resource management for such clouds stands in trying to be locality-aware: for the needs of an application, several virtual machines may exchange data. Placing them *close* to each others can significantly improve the performance of the application they compose. More generally, building an overlay network which takes the hierarchical aspects of the platform without being a hierarchical overlay – which comes with load balancing and resilience issues is a challenge by itself.

The results of these works are planned to be integrated into the Discovery initiative [52] which aims at revisiting OpenStack to offer a cloud stack able to manage utility computing platforms where computing resources are located in small computing centers in the backbone's PoPs (Point of Presence) and interconnected through the backbone's internal links.

3.2.2. Resource management in mobile edge clouds

Mobile edge cloud (MEC) infrastructures are composed of compute, storage and networking resources located at the edge of wide-area networks, in immediate proximity to the end users. Instead of treating the mobile operator's network as a high-latency dumb pipe between the end users and the external service providers, MEC platforms aim at deploying cloud functionalities *within* the mobile phone network, inside or close to the mobile access points. Doing so is expected to deliver added value to the content providers and the end users by enabling new types of applications ranging from Internet-of-Things applications to extremely interactive systems (e.g., augmented reality). Simultaneously, it will generate extra revenue streams for the mobile network operators, by allowing them to position themselves as cloud computing operators and to rent their already-deployed infrastructure to content and application providers.

Mobile edge clouds have very different geographical distribution compared to traditional clouds. While traditional clouds are composed of many reliable and powerful machines located in a very small number of data centers and interconnected by very high-speed networks, mobile edge cloud are composed of a very large number of points-of-presence with a couple of weak and potentially unreliable servers, interconnected with each other by commodity long-distance networks. This creates new demands for the organization of a scalable mobile edge computing infrastructure, and opens new directions for research.

The main challenges that we plan to address are:

- How should an edge cloud infrastructure be designed such that it remains scalable, fault-tolerant, controllable, energy-efficient, etc.?
- How should applications making use of edge clouds be organized? One promising direction is to explore the extent to which stream-data processing platforms such as Apache Spark and Apache Flink can be adapted to become one of the main application programming paradigms in such environments.

3.2.3. *Self-optimizing applications in multi-cloud environments*

As the use of cloud computing becomes pervasive, the ability to deploy an application on a multi-cloud infrastructure becomes increasingly important. Potential benefits include avoiding dependence on a single vendor, taking advantage of lower resource prices or resource proximity, and enhancing application availability. Supporting multi-cloud application management involves two tasks. First, it involves selecting an initial multi-cloud application deployment that best satisfies application objectives and optimizes performance and cost. Second, it involves dynamically adapting the application deployment in order to react to changes in execution conditions, application objectives, cloud provider offerings, or resource prices. Handling price changes in particular is becoming increasingly complex. The reason is the growing trend of providers offering sophisticated, dynamic pricing models that allow buying and selling resources of finer granularities for shorter time durations with varying prices.

Although multi-cloud platforms are starting to emerge, these platforms impose a considerable amount of effort on developers and operations engineers, provide no support for dynamic pricing, and lack the responsiveness and scalability necessary for handling highly-distributed, dynamic applications with strict quality requirements. The goal of this work is to develop techniques and mechanisms for automating application management, enabling applications to cope with and take advantage of the dynamic, diverse, multi-cloud environment in which they operate.

The main challenges arising in this context are:

- selecting effective decision-making approaches for application adaptation,
- supporting scalable monitoring and adaptation across multiple clouds,
- performing adaptation actions in a cost-efficient and safe manner.

3.3. Greening clouds

ICT (Information and Communications Technologies) ecosystem now approaches 5% of world electricity consumption and this ICT energy use will continue grow fast because of the information appetite of Big Data, big networks and big infrastructures as Clouds that unavoidably leads to big power.

3.3.1. *Smart grids and clouds*

We propose exploiting Smart Grid technologies to come to the rescue of energy-hungry Clouds. Unlike in traditional electrical distribution networks, where power can only be moved and scheduled in very limited ways, Smart Grids dynamically and effectively adapt supply to demand and limit electricity losses (currently 10% of produced energy is lost during transmission and distribution).

For instance, when a user submits a Cloud request (such as a Google search for instance), it is routed to a data center that processes it, computes the answer and sends it back to the user. Google owns several data centers spread across the world and for performance reasons, the center answering the user's request is more likely to be the one closest to the user. However, this data center may be less energy efficient. This request may have consumed less energy, or a different kind of energy (renewable or not), if it had been sent to this further data center. In this case, the response time would have been increased but maybe not noticeably: a different trade-off between quality of service (QoS) and energy-efficiency could have been adopted.

While Clouds come naturally to the rescue of Smart Grids for dealing with this big data issue, little attention has been paid to the benefits that Smart Grids could bring to distributed Clouds. To our knowledge, no previous work has exploited the Smart Grids potential to obtain and control the energy consumption of entire Cloud infrastructures from underlying facilities such as air conditioning equipment (which accounts for 30% to 50% of a data center's electricity bill) to network resources (which are often operated by several actors) and to computing resources (with their heterogeneity and distribution across multiple data centers). We aim at taking advantage of the opportunity brought by the Smart Grids to exploit renewable energy availability and to optimize energy management in distributed Clouds.

3.3.2. Energy cost models

Cloud computing allows users to outsource the computer resources required for their applications instead of using a local installation. It offers on-demand access to the resources through the Internet with a pay-as-you-go pricing model.However, this model hides the electricity cost of running these infrastructures.

The costs of current data centers are mostly driven by their energy consumption (specifically by the air conditioning, computing and networking infrastructure). Yet, current pricing models are usually static and rarely consider the facilities' energy consumption per user. The challenge is to provide a fair and predictable model to attribute the overall energy costs per virtual machine and to increase energy-awareness of users.

Another goal consists in better understanding the energy consumption of computing and networking resources of Clouds in order to provide energy cost models for the entire infrastructure including incentivizing cost models for both Cloud providers and energy suppliers. These models will be based on experimental measurement campaigns on heterogeneous devices. Inferring a cost model from energy measurements is an arduous task since simple models are not convincing, as shown in our previous work. We aim at proposing and validating energy cost models for the heterogeneous Cloud infrastructures in one hand, and the energy distribution grid on the other hand. These models will be integrated into simulation frameworks in order to validate our energy-efficient algorithms at larger scale.

3.3.3. Energy-aware users

In a Cloud moderately loaded, some servers may be turned off when not used for energy saving purpose. Cloud providers can apply resource management strategies to favor idle servers. Some of the existing solutions propose mechanisms to optimize VM scheduling in the Cloud. A common solution is to consolidate the mapping of the VMs in the Cloud by grouping them in a fewer number of servers. The unused servers can then be turned off in order to lower the global electricity consumption.

Indeed, current work focuses on possible levers at the virtual machine suppliers and/or services. However, users are not involved in the choice of using these levers while significant energy savings could be achieved with their help. For example, they might agree to delay slightly the calculation of the response to their applications on the Cloud or accept that it is supported by a remote data center, to save energy or wait for the availability of renewable energy. The VMs are black boxes from the Cloud provider point of view. So, the user is the only one to know the applications running on her VMs.

We plan to explore possible collaborations between virtual machine suppliers, service providers and users of Clouds in order to provide users with ways of participating in the reduction of the Clouds energy consumption. This work will follow two directions: 1) to investigate compromises between power and performance/service quality that cloud providers can offer to their users and to propose them a variety of options adapted to their workload; and 2) to develop mechanisms for each layer of the Cloud software stack to provide users with a quantification of the energy consumed by each of their options as an incentive to become greener.

3.4. Securing clouds

3.4.1. Security monitoring SLO

While the trend for companies to outsource their information system in clouds is confirmed, the problem of securing an information system becomes more difficult. Indeed, in the case of infrastructure clouds, physical resources are shared between companies (also called tenants) but each tenant controls only parts of the shared resources, and, thanks to virtualization, the information system can be dynamically and automatically reconfigured with added or removed resources (for example starting or stopping virtual machines), or even moved between physical resources (for example using virtual machine migration). Partial control of shared resources brings new classes of attacks between tenants, and security monitoring mechanisms to detect such attacks are better placed out of the tenant-controlled virtual information systems, that is under control of the cloud provider. Dynamic and automatic reconfigurations of the information system make it unfeasible for a tenant's security administrator to setup the security monitoring components to detect attacks, and thus an automated self-adaptable security monitoring service is required.

Combining the two previous statements, there is a need for a dependable, automatic security monitoring service provided to tenants by the cloud provider. Our goal is to address the following challenges to design such a security monitoring service:

1. to define relevant Service-Level Objectives (SLOs) of a security monitoring service, that can figure in the Service-Level Agreement (SLA) signed between a cloud provider and a tenant;
2. to design heuristics to automatically configure provider-controlled security monitoring software components and devices so that SLOs are reached, even during automatic reconfigurations of tenants' information systems;
3. to design evaluation methods for tenants to check that SLOs are reached.

Moreover in challenges 2 and 3 the following sub-challenges must be addressed:

- although SLAs are bi-lateral contracts between the provider and each tenant, the implementation of the contracts is based on shared resources, and thus we must study methods to combine the SLOs;
- the designed methods should have a minimal impact on performance.

3.4.2. Data Protection in Cloud-based IoT Services

The Internet of Things is becoming a reality. Individuals have their own swarm of connected devices (e.g. smartphone, wearables, and home connected objects) continually collecting personal data. A novel generation of services is emerging exploiting data streams produced by the devices' sensors. People are deprived of control of their personal data as they don't know precisely what data are collected by service providers operating on Internet (oISP), for which purpose they could be used, for how long they are stored, and to whom they are disclosed. In response to privacy concerns the European Union has introduced, with the Global Data Protection Regulation (GDPR), new rules aimed at enforcing the people's rights to personal data protection. The GDPR also gives strong incentives to oISPs to comply. However, today, oISPs can't make their systems GDPR-compliant since they don't have the required technologies. We argue that a new generation of system is mandatory for enabling oISPs to conform to the GDPR. We plan to to design an open source distributed operating system for native implementation of new GDPR rules and ease the programming of compliant cloud-based IoT services. Among the new rules, transparency, right of erasure, and accountability are the most challenging ones to be implemented in IoT environments but could fundamentally increase people's confidence in oISPs. Deployed on individuals' swarms of devices and oISPs' cloud-hosted servers, it will enforce detailed data protection agreements and accountability of oISPs' data processing activities. Ultimately we will show to what extend the new GDPR rules can be implemented for cloud-based IoT services.

3.5. Experimenting with Clouds

Cloud platforms are challenging to evaluate and study with a sound scientific methodology. As with any distributed platform, it is very difficult to gather a global and precise view of the system state. Experiments are not reproducible by default since these systems are shared between several stakeholder. This is even worsen by the fact that microscopic differences in the experimental conditions can lead to drastic changes since typical Cloud applications continuously adapt their behavior to the system conditions.

3.5.1. Experimentation methodologies for clouds

We propose to combine two complementary experimental approaches: direct execution on testbeds such as Grid'5000, that are eminently believable but rather labor intensive, and simulations (using *e.g.* SimGrid) that are much more light-weighted, but requires are careful assessment. One specificity of the Myriads team is that we are working on these experimental methodologies *per se*, raising the standards of *good experiments* in our community.

We plan to make SimGrid widely usable beyond research laboratories, in order to evaluate industrial systems and to teach the future generations of cloud practitioners. This requires to frame the specific concepts of Cloud systems and platforms in actionable interfaces. The challenge is to make the framework both easy to use for simple studies in educational settings while modular and extensible to suit the specific needs of every advanced industrial-class users.

We aim at leveraging the convergence opportunities between methodologies by further bridging simulation and real testbeds. The predictions obtained from the simulator should be validated against some real-world experiments obtained on the target production platform, or on a similar platform. This (in)validation of the predicted results often improves the understanding of the modeled system. On the other side, it may even happen that the measured discrepancies are due to some mis-configuration of the real platform that would have been undetected without this (in)validation study. In that sense, the simulator constitutes a precious tool for the quality assurance of real testbeds such as Grid'5000.

Scientists need more help to make there Cloud experiments fully reproducible, in the sprit of Open Science exemplified by the HAL Open Archive, actively backed by Inria. Users still need practical solutions to archive, share and compare the whole experimental settings, including the raw data production (particularly in the case of real testbeds) and their statistical analysis. This is a long lasting task to which we plan to collaborate through the research communities gathered around the Grid'5000 and SimGrid scientific instruments.

Finally, since correction and performance can constitute contradictory goals, it is particularly important to study them jointly. To that extend, we want to bridge the performance studies, that constitute our main scientific heritage, to correction studies leveraging formal techniques. SimGrid already includes to exhaustively explore the possible executions. We plan to continue this work to ease the use of the relevant formal methods to the experimenter studying Cloud systems.

3.5.2. Use cases

In system research it is important to work on real-world use cases from which we extract requirements inspiring new research directions and with which we can validate the system services and mechanisms we propose. In the framework of our close collaboration with the Data Science Technology department of the LBNL, we will investigate cloud usage for scientific data management. Next-generation scientific discoveries are at the boundaries of datasets, e.g., across multiple science disciplines, institutions and spatial and temporal scales. Today, data integration processes and methods are largely adhoc or manual. A generalized resource infrastructure that integrates knowledge of the data and the processing tasks being performed by the user in the context of the data and resource lifecycle is needed. Clouds provide an important infrastructure platform that can be leveraged by including knowledge for distributed data integration.

4. Application Domains

4.1. Application Domains

The Myriads team investigates the design and implementation of system services. Thus its research activities address a broad range of application domains. We validate our research results with selected use cases in the following application domains:

- Web services, Service oriented applications,
- Business applications,
- Bio-informatics applications,
- Computational science applications,
- Data science applications,
- Numerical simulations,
- Energy and sustainable development,
- Smart cities.

5. Highlights of the Year

5.1. Highlights of the Year

- The PaaSage European project was successfully completed in November 2016 with an excellent rating from the reviewers. The PaaSage project developed a model-based, cross-cloud development and deployment platform that overcomes platform heterogeneity while enabling dynamic, fully-automated application scaling and cloud bursting. The main Myriads contribution is the Adapter subsystem, responsible for supporting dynamic, cross-cloud application adaptation.

5.1.1. Awards

- Baptiste Goupille-Lescar won the prize of the organizing committee of MMS Challenge 2016 (INSA Science Day).
- Anna Giannakou won the "Most Promising Experiment" award at the Grid'5000 winter school in February 2016 for her work "Towards Self Adaptable Security Monitoring in IaaS Clouds".

6. New Software and Platforms

6.1. ConPaaS

KEYWORDS: Cloud computing - PaaS
SCIENTIFIC DESCRIPTION ConPaaS [60] is a runtime environment for hosting applications in the cloud. It aims at offering the full power of the cloud to application developers while shielding them from the associated complexity of the cloud. ConPaaS is designed to host both high-performance scientific applications and online Web applications. It automates the entire life-cycle of an application, including collaborative development, deployment, performance monitoring, and automatic scaling. This allows developers to focus their attention on application-specific concerns rather than on cloud-specific details.
FUNCTIONAL DESCRIPTION

ConPaaS is a runtime environment for hosting applications in the cloud. It aims at offering the full power of the cloud to application developers while shielding them from the associated complexity of the cloud. ConPaaS is designed to host both high-performance scientific applications and online Web applications. It automates the entire life-cycle of an application, including collaborative development, deployment, performance monitoring, and automatic scaling. This allows developers to focus their attention on application-specific concerns rather than on cloud-specific details.

- Participants: Guillaume Pierre, Eliya Buyukkaya, Ancuta Iordache, Morteza Neishaboori, Dzenan Softic, Genc Tato and Teodor Crivat
- Contact: Guillaume Pierre
- URL: http://www.conpaas.eu/

6.2. GinFlow

KEYWORDS: Workflow - Distributed computing - Distributed - Distributed Applications - Framework
FUNCTIONAL DESCRIPTION GinFlow decentralizes the coordination of the execution of workflow-based applications. GinFlow relyies on an architecture where multiple service agents (SA) coordinate each others through a shared space containing the workflow description and current status. GinFlow allows the user to define several variants of a workflow and to switch from one to the other during run time.

- Participants: Matthieu Simonin, Cédric Tedeschi, Hector Fernandez, Javier Rojas Balderrama and Thierry Priol
- Partner: Université de Rennes 1
- Contact: Cédric Tedeschi
- URL: http://ginflow.inria.fr

6.3. Merkat

KEYWORDS: Resource management - Cloud - Elastic scaling - Market mechanisms - Service Level Objectives - HPC
FUNCTIONAL DESCRIPTION

Merkat is a platform that allows users of an organization to automatically manage and scale their applications while maximizing the infrastructure's utilization [12]. Merkat is generic and extensible, allowing users to automate the application deployment and management process. Users have the flexibility to control how many resources are allocated to their applications and to define their own resource demand adaptation policies. Merkat applies an unique approach to multiplex the infrastructure capacity between the applications, by implementing a proportional-share market and allowing applications to adapt autonomously to resource price and their given performance objectives. The price of the acquired resources acts as a control mechanism to ensure that resources are distributed to applications according to the user's value for them. Merkat was evaluated on Grid'5000 with several scientific applications.

- Participants: Stefania Costache, Christine Morin and Nikos Parlavantzas
- Contact: Nikos Parlavantzas
- URL: http://www.irisa.fr/myriads/software/Merkat/

6.4. Meryn

KEYWORDS: Resource management - PaaS - Cloud - Market mechanisms - Service Level Agreements
FUNCTIONAL DESCRIPTION

Meryn is an open, SLA-driven PaaS architecture that supports cloud bursting and allows hosting an extensible set of application types. Meryn relies on a decentralized optimization policy that aims at maximizing the overall provider profit, taking into account the penalties incurred when quality guarantees are unsatisfied. Meryn builds on the Snooze VM manager, and currently supports batch and MapReduce applications.

- Participants: Nikos Parlavantzas, Djawida Dib and Christine Morin
- Contact: Nikos Parlavantzas
- URL: http://www.irisa.fr/myriads/software/Meryn/

6.5. PaaSage Adapter

KEYWORDS: Cloud computing - Dynamic adaptation - Cloud applications management
FUNCTIONAL DESCRIPTION

The purpose of the Adapter is to transform the current configuration of a cloud application into a target configuration in an efficient and safe way. The Adapter is part of PaaSage, an open-source platform for modeling, deploying and executing applications on different clouds in an optimal manner. The Adapter has the following responsibilities: (1) validating reconfiguration plans, (2) applying the plans to the running system, and (3) maintaining an up-to-date representation of the current system state.

- Participants: Nikos Parlavantzas, Arnab Sinha, Manh Linh Pham, and Christine Morin
- Contact: Nikos Parlavantzas
- URL: https://team.inria.fr/myriads/software-and-platforms/paasage-adapter/

6.6. Resilin

KEYWORDS: Map Reduce - Parallel processing - Hadoop - Cloud - PaaS
FUNCTIONAL DESCRIPTION

Resilin is an open-source system for creating and managing MapReduce execution platforms over clouds. Resilin is compatible with the Amazon Elastic MapReduce (EMR) API, but it goes beyond Amazon's proprietary EMR solution in allowing users (e.g. companies, scientists) to leverage resources from one or more public and/or private clouds. This enables performing MapReduce computations over a large number of geographically-distributed and diverse resources. Resilin can be deployed across most of the open-source and commercial IaaS cloud management systems (e.g., OpenStack, OpenNebula, Amazon EC2). Once deployed, Resilin takes care of provisioning Hadoop clusters and submitting MapReduce jobs, allowing users to focus on writing their MapReduce applications rather than managing cloud resources. Resilin is implemented in the Python language and uses the Apache Libcloud library to interact with IaaS clouds. Resilin has been evaluated on multiple clusters of the Grid'5000 experimentation testbed. The results show that Resilin enables the use of geographically distributed resources with a limited impact on MapReduce job execution time.

- Participants: Ancuta Iodache, Christine Morin, Pierre Riteau, Nikos Parlavantzas and Matthieu Simonin
- Contact: Christine Morin
- URL: http://resilin.inria.fr

6.7. SAIDS

KEYWORDS: Cloud - Security
FUNCTIONAL DESCRIPTION SAIDS is a self-adaptable intrusion detection system for IaaS clouds. To maintain an effective level of intrusion detection, SAIDS monitors changes in the virtual infrastructure of a Cloud environment and reconfigures its components (security probes) accordingly. SAIDS can also reconfigure probes in the case of a change in the list of running services.

- Participants: Anna Giannakou, Jean-Léon Cusinato, Christine Morin, Jean-Louis Pazat, Louis Rilling and Fergal Martin-Tricot,
- Contact: Christine Morin
- URL: https://bil.inria.fr

6.8. SimGrid

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

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

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

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

- Participants: Martin Quinson,
- Partners: Frédéric Suter, Arnaud Legrand, Adrien Lèbre, Luka Stanisic, Augustin Degomme.
- Contact: Martin Quinson
- URL: http://simgrid.gforge.inria.fr/

6.9. Snooze

KEYWORDS: Energy management - Fault-tolerance - Self-organization - Self-healing - Cloud computing - Consolidation - Virtualization

SCIENTIFIC DESCRIPTION

Snooze is a scalable, resilient and energy-aware virtual machine management framework for clouds. It is the result of Eugen Feller's PhD thesis which has been funded by the ANR EcoGrappe project (2008 - 2012).

The objectives of the Snooze ADT are threefold: (i) to distribute Snooze system as an open source software and to provide support to the user community (ii) to implement additional features to make it more user-friendly (iii) to integrate it with other open source software stacks and in public testbeds to favour its dissemination.

Snooze is a highly modular system for IaaS clouds. For the scientific community, it constitutes a unique framework for the experimentation of resource management policies in a real system. More generally, it allows any organization to operate a large-scale cluster as a resilient and energy-aware computing infrastructure enabling on demand provisioning of virtual clusters.

FUNCTIONAL DESCRIPTION

Snooze is a self-organizing and energy aware Cloud management framework.

Snooze is a novel Infrastructure-as-a-Service (IaaS) cloud-management system, which is designed to scale across many thousands of servers and virtual machines (VMs) while being easy to configure, highly available, and energy efficient. For scalability, Snooze performs distributed VM management based on a hierarchical architecture. To support ease of configuration and high availability Snooze implements self-configuring and self-healing features. Finally, for energy efficiency, Snooze integrates a holistic energy management approach via VM resource (i.e. CPU, memory, network) utilization monitoring, underload/overload detection and mitigation, VM consolidation (by implementing a modified version of the Sercon algorithm [59]), and power management to transition idle servers into a power saving mode. Snooze is a highly modular piece of software. It has been extensively evaluated on the Grid'5000 testbed using realistic applications.

- Participants: Eugen Feller, Christine Morin, Jiajun Cao, Gene Cooperman, Yvon Jégou, David Margery and Matthieu Simonin

- Contact: Christine Morin
- URL: http://snooze.inria.fr/

6.10. VEP

KEYWORDS: Cloud - Security - Computing - IaaS - Standards - OVF - CIMI - SLA

SCIENTIFIC DESCRIPTION

Virtual Execution Platform (VEP) is a Contrail service that sits just above IaaS layer at the service provider end of the Contrail cloud federation. The VEP service provides a uniform interface for managing the whole lifecycle of elastic applications on the cloud and hides the details of the IaaS layer to the user. VEP applications are described in OVF (Open Virtualization Format) standard format. Resource usage is controlled by CEE (Constrained Execution Environment) rules which can be derived from SLAs (Service Level Agreement). The VEP service integrates a monitoring system where the major events about the application, mainly resource usage, are made available to the user.

The VEP service provides a RESTful interface and can be exploited directly by users on top of the provider IaaS. OpenNebula and OpenStack IaaS frameworks were initially supported. During the VEP-S EIT ICT Labs activity in 2014, VEP was extended with a new OCCI IaaS driver which allows to control any IaaS framework providing a standard OCCI API. Support for the new OCCI SLA proposition from OGF has also been added and allows to represent the VEP CEEs in a standard format. Finally, during this activity, the Zabbix open source distributed monitoring system was integrated to VEP.

FUNCTIONAL DESCRIPTION VEP is a management tool for IaaS clouds with a REST interface and simple GUI for administrator. It is an extensible and reusable software for easy deployment of distributed applications. It provides advance reservation, pro-active fault tolerance. It is SLA aware and manages elasticity.

- Participants: Yvon Jégou, Roberto Gioacchino Cascella, Florian Dudouet, Filippo Gaudenzi, Christine Morin and Arnab Sinha
- Contact: Christine Morin
- URL: https://project.inria.fr/vep/

7. New Results

7.1. Scaling Clouds

7.1.1. Heterogeneous Resource Management

Participants: Baptiste Goupille-Lescar, Ancuta Iordache, Christine Morin, Manh Linh Pham, Nikos Parlavantzas, Guillaume Pierre, Arnab Sinha.

7.1.1.1. High performance in the cloud with FPGA virtualization

Participants: Ancuta Iordache, Guillaume Pierre.

Cloud platforms are becoming increasingly heterogeneous, with the availability of large numbers of virtual machine instance types as well as accelerator devices such as GPUs. In collaboration with Maxeler technologies, we have proposed a technique to virtualize FPGAs and make them available as first-class high-performance computation devices in the cloud [24]. The increasing variety of computation, storage and networking resources in the cloud is an opportunity for adjusting the provisioned resources to the individual needs of each application, but making an informed choice is extremely difficult. We therefore proposed application profiling techniques which can automatically identify the configuration which provides the best performance/cost tradeoff [49]. These two results were developed as part of the HARNESS European project, and they constitute Anca Iordache's PhD thesis [50]. FPGA virtualization is being further developed by Maxeler technologies toward commercial exploitation, and application profiling has been integrated in the open-source ConPaaS platform.

7.1.1.2. Multi-cloud application execution

Participants: Manh Linh Pham, Nikos Parlavantzas, Arnab Sinha.

Within the PaaSage European project, we improved and extended the Adapter subsystem, the part of the PaaSage platform that dynamically adapts the application deployment to changes in current runtime conditions [45]. Specifically, we added full support for causal connection between the running system and the runtime model and extended the plan validation functionality to use historical reconfiguration information. Moreover, we assisted industrial PaaSage partners with applying the PaaSage platform in diverse business scenarios.

7.1.1.3. Adaptive resource management for high-performance, multi-sensor systems

Participants: Baptiste Goupille-Lescar, Christine Morin, Nikos Parlavantzas.

In the context of our collaboration with Thales Research and Technology, we are applying cloud resource management techniques to high-performance, multi-sensor, embedded systems with real-time constraints. The objective is to increase the flexibility and efficiency of resource allocation in such systems, enabling the execution of dynamic sets of applications with strict QoS requirements. In 2016, we focused on characterising the targeted applications and platforms and developing a simulator in order to explore relevant resource management solutions. This work is performed in the context of Baptiste Goupille-Lescar's PhD work.

7.1.2. Distributed Cloud Computing

Participants: Nikos Parlavantzas, Jean-Louis Pazat, Guillaume Pierre, Genc Tato, Cédric Tedeschi, Alexandre Van Kempen.

7.1.2.1. Application self-optimization in multi-cloud environments

Participant: Nikos Parlavantzas.

Current approaches to application adaptation in multi-cloud environments are typically static, platform dependent, complex, and error prone. To address these limitations, we are combining the use of software product lines (SPLs) with models@run-time techniques. This work is performed in the context of the thesis of Carlos Ruiz Diaz, a PhD student at the University of Guadalajara, co-advised by Nikos Parlavantzas. The work focuses on the development of an SPL-based framework supporting initial cloud configuration as well as proactive, dynamic adaptation in a systematic, platform-independent way. The evaluation of this framework is currently in progress.

7.1.2.2. Edge clouds

Participants: Guillaume Pierre, Genc Tato, Cédric Tedeschi, Alexandre Van Kempen.

Mobile edge cloud computing aims to deploy cloud resources even closer to the end users, typically within mobile network access points. This is useful for hyper-interactive applications such as augmented reality which demand ultra-low network latencies (2-5 ms) between the end-user device and the cloud instances serving it. In contrast, current mobile networks exhibit network latencies in the order of 50-150 ms between the device and any cloud. We extended the ConPaaS open-source cloud platform to support the deployment of cloud applications in a distributed set of Raspberry Pi machines: instead of reaching the cloud through a wide-area network, in this setup each cloud node is also equipped with a wifi hotspot which allows local users to access it directly [53]. This work is ongoing, and a paper on this topic is currently being reviewed.

Getting closer to the edge user can be done through provisioning computing resources in Points of Presence (PoPs) within the telco's backbone network. The Discovery project [52] aims at revisiting the OpenStack Cloud stack to allow to disperse several smaller cloud facilities and connect them together to make them appear as a single Cloud entity. Genc Tato's PhD aims at proposing the building blocks on top of such an infrastructure to abstract out the network, route queries, store and retrieve objects (VMs and data). We have devised an overlay network to support such functionalities keeping in mind to maximise the laziness of the maintenance protocol to avoid any useless cost. A paper is being written on the subject.

7.1.2.3. Community Clouds

Participant: Jean-Louis Pazat.

Hosting services on an edge infrastructure based on devices owned and operated by end-users may be interesting for serving a community of users. However, these devices (such as internet boxes, disks or small computers have heterogeneous capabilities and no guaranteed availability. It is therefore challenging to ensure to the guest application a minimal hosting service level, like availability or Quality of Service. The management of the hosting service should adapt to the characteristics of the infrastructure. We are designing an architecture for a middleware capable of adapting the deployment of services on edge devices to ensure a given Quality of Service to access the service. While the middleware requires a minimal knowledge of the underlying infrastructure, its adaptation decisions are based on the feedbacks of users of the deployed service, like measured network latency. The environment relies on the use of micro-services which are composed to build the end-user services. This allows many adaptation strategies to adapt the system during run-time.

7.1.3. Scaling workflows with GinFlow

Participants: Matthieu Simonin, Cédric Tedeschi.

In 2016, we deployed GinFlow over 800 cores of the Grid'5000 platform, running Montage workflows comprising 118 tasks, and artificial workflows made of more than 3000 tasks. The ability of GinFlow to support adaptation and versioning of workflow with seamless transitions between workflow alternatives at runtime has been validated experimentally and presented on the Inria booth at SuperComputing in November 2016. These results have been presented at the IPDPS conference [32], and have been submitted to a journal special issue on workflows.

7.2. Greening Clouds

7.2.1. Energy Models

Participants: Yvon Jégou, Anne-Cécile Orgerie, Edouard Outin, Jean-Louis Pazat, Martin Quinson.

Simulating the impact of DVFS within SimGrid 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. The goal of this ongoing work is to enable energy-aware experimentation within the SimGrid simulation toolkit, by introducing a model of energy consumption for computing applications making use of Dynamic Voltage and Frequency Scaling (DVFS) techniques.

Simulating Energy Consumption of Wired Networks 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. This simulator embeds green levers: low power idle (IEEE 802.3az) and adaptive link rate. An article is currently under review on this topic.

Multicriteria scheduling for large-scale HPC environments Energy consumption is one of the main limiting factor for the design and deployment of large scale numerical infrastructures. The road towards "Sustainable Exascale" is a challenge with a target of 50 Gflops per watt. As platforms become more and more heterogeneous (co-processors, GPUs, low power processors...), an efficient scheduling of applications and services at large scale remains a challenge. In this context, we explore a multicriteria scheduling model and framework for large scale HPC systems. This work is done in collaboration with ROMA and Avalon teams from LIP in Lyon [29], [37].

Dynamic resource management for energy-efficiency The B-Com project, a joint private/public focusing on transfer, targets the design and the implementation of Watcher, a software module used to optimize an OpenStack cloud (in terms of performance, storage optimization or energy savings). This Software module is in the "Big Tent" software development process of OpenStack. In cooperation with Olivier Barais (Diverse Inria Team), we focus on dynamic management of cloud resources for energy-efficiency. Our approach relies on machine learning techniques, models@run-time and dynamic adaptation, and is intended to be included in Watcher. At regular intervals of time, we optimize the use of cloud resources by checking if a better placement of Virtual Machines on physical resources can be achieved, taking into account the migration cost. To achieve this, we have an energy model of the resources which is regularly updated using machine learning techniques that helps optimization algorithms to check if a better configuration can be reached energy-wise. This year we worked on the evaluation of the energy model [28].

7.2.2. Involving users in Energy Saving

Participants: Deborah Agarwal, Ismael Cuadrado Cordero, David Guyon, Christine Morin, Anne-Cécile Orgerie.

Energy-efficient cloud elasticity for data-driven applications Data centers hosting cloud systems consume enormous amounts of energy. Reducing this consumption becomes an urgent challenge with the rapid growth of cloud utilization. An existing solution to lower this consumption is to turn off as many servers as possible, but these solutions do not involve the user as a main lever to save energy. We introduce a system that proposes to the user to run her application with degraded performance in order to promote a better consolidation and thus to turn off more servers. Experimentation results using the Montage workflow show promising outcomes [47], [48]. We also performed a simulation-based evaluation on how much an energy-aware cloud system could save in energy consumed depending on the proportion of users selecting a green execution mode. These results based on the simulation of two typical daily uses of a data center running 3 real scientific applications will be published in Euromicro PDP 2017.

Energy-efficient and network-aware resource allocation in Cloud infrastructures The ever-growing appetite of new applications for network resources leads to an unprecedented electricity bill, and for these bandwidth-hungry applications, networks can become a significant bottleneck. Towards this end, we proposed microclouds, a fully autonomous energy-efficient subnetwork of clients of the same service, designed to keep the greenest path between its node. This semi-decentralized PaaS architecture for real-time multiple-users applications geographically distributes the computation among the clients of the cloud, moving the computation away from the datacenter to save energy - by shutting down or downgrading non utilized resources such as routers and switches, servers, etc. - and provides lower latencies for users. In this work, we have also analyzed the use of incentives for Mobile Clouds, and proposed a new auction system adapted to the high dynamism and heterogeneity of these systems [20], [19] [46].

7.2.3. Exploiting Renewable Energy in Datacenters

Participants: Sabbir Hasan Rochi, Yunbo Li, Anne-Cécile Orgerie, Jean-Louis Pazat.

Resource allocation in a Cloud partially powered by renewable energy sources We propose here to design a disruptive approach to Cloud resource management which takes advantage of renewable energy availability to perform opportunistic tasks. 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 integrates a prediction model to be able to forecast these extra-power periods of time in order to schedule more work during these periods. This work is done in collaboration with Ascola team from LINA in Nantes [44], [51][9].

Creating green-energy adaptivity awareness in SaaS application In addition to “green” resource allocation at the IaaS level in Datacenters, we think that users should be involved in “greening” their energy use (SaaS level). We propose that applications should have multiple “modes” of execution, each mode using a different level of energy and providing a different service level. For example, a B2C application may provide more or less recommandations . If this applcation can be dynamically swiched between these modes depending on the availability of green energy, the IaaS can optimize resource allocation better. To enforce his, we have designed green energy aware controllers.

This work is done in collaboration with Ascola team [23], [9].

7.3. Securing Clouds

7.3.1. Security monitoring in clouds

Participants: Jean Leon Cusinato, Anna Giannakou, Fergal Martin-Tricot, Christine Morin, Jean-Louis Pazat, Louis Rilling, Amir Teshome Wonjiga.

In the INDIC project we aim at making security monitoring a dependable service for IaaS cloud customers. To this end, we study three topics:

- defining relevant SLA terms for security monitoring,
- enforcing and verifying SLA terms,
- making the SLA terms enforcement mechanisms self-adaptable to cope with the dynamic nature of clouds.

The considered enforcement and verification mechanisms should have a minimal impact on performance.

In 2016 we improved the SAIDS approach, that we proposed in 2015, and that makes a network intrusion detection system (NIDS) deployed in a cloud operator infrastructure self-adaptable. In particular, we validated that the approach is generic enough to handle signature-based NIDSs (support for Snort and Suricata was implemented) as well as event-based NIDSs (support for Bro was implemented). An experimental evaluation of SAIDS has also been started in order to submit a full paper for publication in 2017. Jean-Léon Cusinato contributed to this work during his master internship.

We also improved the AL-SAFE approach, that we proposed in 2015, and that secures an application-level firewall by isolating it from the customer virtual machine and makes it self-adaptable [36], [35]. In particular, we validated that the self-adaptation architecture introduced for SAIDS could be reused to address firewalls, and the prototype was improved to implement stateful filtering. Fergal Martin-Tricot contributed to this work during his master internship. We also evaluated AL-SAFE experimentally on the prototype as well as analytically regarding the security correctness. The design and the evaluation of AL-SAFE were published in the CloudCom 2016 conference [21].

Regarding SLA definition and enforcement, in 2016 we have studied a verification method to enable a Cloud customer to verify that an NIDS located in the operator infrastructure is configured correctly according to the Service-Level Objectives (SLO) figuring in the SLA. A simple example of SLO is being used for this study, and further work should address more complete SLO regarding NIDSs. A prototype of the proposed verification method was implemented on OpenStack and Open vSwitch, and the NIDS software used is Snort. An evaluation of the verification method has been started and will include both experiments on the Grid'5000 platform and a correctness analysis. The design and evaluation of the verification method will be submitted in a full paper for publication in 2017.

7.3.2. Risk assessment in clouds

Participant: Christine Morin.

Attack graphs are leveraged in networks to exhibit the various scenarios available to compromise the system. They allow to uncover vulnerabilities chains exploitable by attackers based on network connectivity and vulnerabilities pre-requisites. In physical infrastructures, the acquisition of the topology has been vastly addressed in existing works with either passive or active discovery methods. Considering the Cloud context, in which virtualization attacks and virtual infrastructure dynamism are introduced, new methods need to be developed. We have designed a topology builder able to keep the topology and connectivity up to date in cloud environments. Based on the use of an IaaS cloud management system and a SDN (Software-Defined

Networking) controller, our approach encompasses two steps: (i) when plugged into a running system, the topology builder retrieves the current topology and builds the associated connectivity: this represents the static topology and connectivity retrieval, in which we assume the network configuration to be fixed ; (ii) the topology builder listens to change events generated inside the infrastructure and within the SDN controller in order to update the topology and connectivity previously built: this represents the dynamic topology and connectivity retrieval. A prototype has been developed based on OpenStack cloud management system and ONOS SDN open source technologies. This work is carried out in the context of Pernelle Mensah's PhD thesis and in collaboration with Nokia and CIDRE Inria project-team.

7.4. Experimenting with Clouds

7.4.1. Simulation

Participants: Simon Bihel, Martin Quinson.

Providing better interfaces to the users for Cloud Studies. Aware that the current user interface is a impediment to the adoption of our framework by the scientific community, we tried to propose a new, simplified API through the internship of Simon Bihel this summer. We identified several use cases and usage scenario that relevant to our context, and started implementing the new interface that we will provide. This work is still under progress.

Production-ready simulator of large-scale distributed systems. We are currently involved in a complete reorganization of the SimGrid implementation. The goal is two-fold: first we want reduce the tool's learning curve to help beginners. At the same time, we want to normalize the tool's internals so that power users can modify it and/or script the kernel behavior easily. Eventually, we are targeting usages in production and teaching contexts. This long term overhaul is still underway.

7.4.2. Experimentation Testbed

Participants: Anirvan Basu, Julien Lefeuvre, David Margery, Pascal Morillon.

Providing ready to use scripts to deploy popular and complex stacks. The study of complex software stacks on Grid'5000 has always been possible due to the reconfigurability properties of the testbed. Nevertheless, for newcomers with little background in system administration, automating the deployment of these stacks on Grid'5000 has always proved difficult. In 2016, we have provided scripts, that users can fork on github to customise to their needs, to deploy OpenStack, Ceph, Hadoop over Ceph or Sparkle. These have been presented to users during the 2016 winter school.

7.4.3. Use cases

Participants: Deborah Agarwal, Yvon Jégou, Nikos Parlavantzas, Manh Linh Pham, Christine Morin, Kartik Sathyanarayanan, Arnab Sinha.

7.4.3.1. Experimental Evaluation of Data Stream Processing Frameworks

We worked on evaluating data stream processing environments deployed in clouds. We compared the throughput, latency and energy consumption of Spark Streaming, Storm and Heron real-time data processing environments executed on top of Linux clusters and on top of virtual clusters deployed on top of the OpenStack IaaS cloud. The preliminary evaluation was conducted using the word count application on the twitter data stream. All experiments were conducted on Grid'5000 experimentation platform. The experimental results are described in a technical report to be published in 2017. This work was carried out by Kartik Sathyanarayanan, a student intern in Myriads team in the framework of DALHIS associate team.

7.4.3.2. Simulation framework for studying between-herd pathogen spread in a region

In our collaboration with Inra in the context of the Mihmes project, we worked on the design of decision tools to evaluate the epidemio-economic effectiveness of disease prevention and control strategies at the scales of the herd, the region and the supply chain. We developed a generic service-based framework to efficiently execute models of infection dynamics in a metapopulation of cattle herds on large-scale computing infrastructures. Our framework has been designed to execute complex regional models combining within-herds epidemiological models. The framework automatically distributes the simulation runs on multiple servers in a cluster and exploits the parallelism of the multicore servers. It relies on OpenMP for parallelizing simulation loops and deals with server heterogeneity and failures. We leveraged PaaSage software stack to deploy the framework on several IaaS clouds.

7.4.3.3. Mobile application for reliable collection of field data for Fluxnet

Critical to the interpretation of Fluxnet carbon flux data is the ancillary information and measurements taken at the tower sites. The submission and update of this data using excel sheets is difficult and error prone. In partnership with ICOS in the framework of DALHIS associate team, we are innovating the data submission and organization method through a responsive web User Interface able to run on desktop, mobile etc.; thus easing the data lookup and entry process from anywhere including the field sites. Continuing with our initial usability feedback experiences gathered last year on the application interface designs, we decided on the mobile application workflow for implementation. We developed a first prototype based on the PhoneGap [0] platform which provided the advantage of the same development code generating mobile application for IOS, Android and Windows platform simultaneously. The main functionality realized in the application prototype is that the user can download all the site data required by logging in through the application; and then view/edit them at the tower site (even in offline mode). The next logical step would be developing the synchronization and validation of data held locally in the application with the servers.

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Contracts with Industry

8.1.1. Technicolor (2016-2017)

Participant: Guillaume Pierre.

Our collaboration with Technicolor focuses on the design of a scalable and elastic virtual vistumer premises equipment based on Network Function Virtualization, Software-Defined Networking and Cloud technologies. In 2016 we completed the system design and an engineer from Technicolor started implementing the system. We expect to conduct further experiments and write a joint publication on this topic in 2017.

8.2. Bilateral Grants with Industry

8.2.1. Thales Research and Technology (2016-2018)

Participants: Baptiste Goupille-Lescar, Christine Morin, Nikos Parlavantzas.

Our collaboration with Thales Research and Technology focuses on the development of distributed Cyber-Physical Systems, such as those developed by Thales to monitor and react to changing physical environments. These systems need to be highly adaptable in order to cope with the dynamism and diversity of their operating environments. Notably, they require distributed, parallel architectures that support dynamic sets of applications, not known in advance, while providing strong QoS guarantees. The objective of this collaboration is to explore adaptive resource management mechanisms for such systems that can adapt to changes in the requirements and in the availability of resources. This contract funds Baptiste Goupille-Lescar's PhD grant.

[0] http://phonegap.com/

8.2.2. Nokia (2015-2018)

Participant: Christine Morin.

Together with CIDRE Inria project-team we are involved in a collaboration with Nokia on security policy adaptation driven by risk evaluation in modern communication infrastructures. To address the need for efficient security supervision mechanisms, approaches such as attack graphs generation, coupled to a risk-based assessment have been used to provide an insight into a system's threat exposure. In comparison to static infrastructures, clouds exhibit a dynamic nature and are exposed to new attack scenarios due to virtualization. The goal of this collaboration is thus to revisit existing methods in the context of clouds. This contract funds Pernelle Mensah's PhD grant. Pernelle is a member of CIDRE project-team.

9. Partnerships and Cooperations

9.1. Regional Initiatives

9.1.1. IRT B-Com

Participants: Yvon Jégou, Edouard Outin, Jean-Louis Pazat.

Yvon Jégou and Jean-Louis Pazat are at IRT B-Com [0] one day per week. With Édouard Outin, B-com PhD student, they contribute to the B-Com *Indeed* project, which aims at developing a distributed cloud software stack with a high degree of adaptability.

In the last period, they were involved in the elaboration of new projects in the Cloud Computing lab of B-Com.

9.1.2. CominLabs EPOC project (2013-2017)

Participants: Sabbir Hasan Rochi, Yunbo Li, Anne-Cécile Orgerie, Jean-Louis Pazat.

In this project, partners aim at focusing on energy-aware task execution from the hardware to application's components in the context of a mono-site data center (all resources are in the same physical location) which is connected to the regular electric Grid and to renewable energy sources (such as windmills or solar cells). In this context, we tackle three major challenges:

- Optimizing the energy consumption of distributed infrastructures and service compositions in the presence of ever more dynamic service applications and ever more stringent availability requirements for services.
- Designing a clever cloud's resource management which takes advantage of renewable energy availability to perform opportunistic tasks, then exploring the trade-off between energy saving and performance aspects in large-scale distributed systems.
- Investigating energy-aware optical ultra high-speed interconnection networks to exchange large volumes of data (VM memory and storage) over very short periods of time.

9.1.3. INDIC - Cybersecurity Pole of Excellence (2014-2018)

Participants: Anna Giannakou, Christine Morin, Jean-Louis Pazat, Louis Rilling, Amir Teshome Wonjiga.

Our study carried out in the framework of a collaboration with DGA-MI aims at defining and enforcing SLA for security monitoring of virtualized information systems. To this aim we study three topics:

- defining relevant SLA terms for security monitoring,
- enforcing and evaluating SLA terms,
- making the SLA terms enforcement mechanisms self-adaptable to cope with the dynamic nature of clouds.

[0] http://b-com.org/wp/

The considered enforcement and evaluation mechanisms should have a minimal impact on performance. The funding from DGA funds two PhD students: Anna Giannakou and Amir Teshome Wonjiga.

9.2. National Initiatives

9.2.1. Inria ADT GinFlow (2014-2016)

Participants: Christine Morin, Matthieu Simonin, Cédric Tedeschi.

The GinFlow technological development action funded by INRIA targets the development of a fully-operational workflow management system based on the HOCL-TS software prototype developed during the PhD thesis of Hector Fernandez between 2009 and 2012. Also, it allows the integration of this software with the TIGRES workflow engine developed at the Lawrence Berkeley National Lab so as to make the workflows submitted using the TIGRES programming model run in a decentralized fashion. These developments led to the release of the GinFlow software and its deposit at the APP (Agence de Protection des Programmes).

9.2.2. Inria ADT SaaP (2016-2018)

Participant: Martin Quinson.

The SaaP technological development action (SimGrid As A Platform) funded by INRIA targets the refactoring of SimGrid to make it ready to use in production and teaching contexts. Our ultimate goal is to sustain the development of the framework by involving 5 to 10 companies that are using it internally. Our target of the teaching context is thus an intermediate goal, as we think that the best solution to ensure the adoption of our tool by the industrial engineers is that they discovered the tool during their studies.

The technical actions envisioned for this ADT are the complete rearchitecturation of the software (to make it easier to script a new model within the tool kernel) and a reorganization of the interfaces (for a better integration in the Java and python language). This work is lead in collaboration with the whole SimGrid community, which provide valuable feedback.

9.2.3. Inria IPL Discovery (2015-2019)

Participants: Anne-Cécile Orgerie, Matthieu Simonin, Cédric Tedeschi.

The Inria IPL Discovery officially started in September 2015. It targets the design, development and deployment of a distributed Cloud infrastructure within the network's backbone. It will be based upon a set of building blocks whose design will take locality as a primary constraint, so as to minimize distant communications and consequently achieve better network traffic, partition management and improved availability.

Its developments are planned to get integrated within the OpenStack framework. An energy/cost benefit analysis of the fully distributed Discovery architecture will also be performed to show the energy efficiency of the chosen approach.

9.2.4. Inria IPL CityLab (2015-2018)

Participant: Christine Morin.

The Inria Project Lab (IPL) CityLab@Inria (https://citylab.inria.fr) studies ICT solutions toward smart cities that promote both social and environmental sustainability. A strong emphasis of the Lab is on the undertaking of a multi-disciplinary research program through the integration of relevant scientific and technology studies, from sensing up to analytics and advanced applications, so as to actually enact the foreseen smart city Systems of Systems. City-scale experiments of the proposed platforms and services are planned in cities in California and France, thereby learning lessons from diverse setups.

Myriads investigates advanced cloud solutions for the Future Internet, which are critical for the processing of urban data. It leverages its experience in cloud computing and Internet of services while expanding its research activities to the design and implementation of cloud services to support crowd-Xing applications and mobile social applications.

In 2016, Christine Morin was involved in the MOOC entitled *Villes Intelligentes : défis technologiques et sociétaux* (Smart cities : technological and social challenges) run on the FUN platform from January to March 2016. She prepared eight sequences on urban data management in clouds. In 2016, we also conducted a comparative experimental evaluation of data stream processing environments executed on clusters and in a cloud. We compared the performance and energy consumption of Heron, Storm and SparkStreaming frameworks.

9.2.5. Inria IPL Hac Specis (2016-2020)

Participants: Anne-Cécile Orgerie, Martin Quinson.

The goal of the HAC SPECIS (High-performance Application and Computers: Studying PErformance and Correctness In Simulation) project (http://hacspecis.gforge.inria.fr/) is to answer methodological needs of HPC application and runtime developers and to allow to study real HPC systems both from the correctness and performance point of view. To this end, we gather experts from the HPC, formal verification and performance evaluation community.

The Anh Pham started a PhD thesis in November 2016, co-advised by Thierry Jéron (team SUMO, formal methods) and Martin Quinson. The envisionned work will pursue the previous efforts to formally assess distributed applications within the SimGrid framework.

9.2.6. COSMIC PRE (2016 - 2018)

Participants: Benjamin Camus, Anne-Cécile Orgerie.

The distributed nature of Cloud infrastructures involves that their components are spread across wide areas, interconnected through different networks, and powered by diverse energy sources and providers, making overall energy monitoring and optimization challenging. The COSMIC project aims at taking advantage of the opportunity brought by the Smart Grids to exploit renewable energy availability and to optimize energy management in distributed Clouds. This PRE, led by Anne-Cécile Orgerie also involves Fanny Dufossé from Dolphin team (Inria Lille) and Benjamin Camus, who has started a 18 months post-doc in October 2016 in the context of this project.

9.2.7. MIHMES ANR Investissements d'Avenir (2012 - 2018)

Participants: Yvon Jégou, Christine Morin, Manh Linh Pham, Nikos Parlavantzas.

The MIMHES project (http://www.inra.fr/mihmes) led by INRA/BioEpAR aims at producing scientific knowledge and methods for the management of endemic infectious animal diseases and veterinary public health risks. Myriads team provides software tools to efficiently manage and ease the use of a distributed computing infrastructure for the execution of different simulation applications.

In 2016, we further developed a distributed framework which allows to exploit multiple compute servers in parallel. Parallelism is exploited both at server level using OpenMP and at data-center level using this framework. To facilitate the deployment of the workloads on heterogeneous environments, this framework limits the requirements concerning the server configurations. They need not share any file system, the workloads can be programmed in differing programming language. These servers need only the capability to communicate through the network. The system allows to dynamically add and stop servers. To some extend, it is tolerant to server failures. The framework had being repackaged to facilitate its reuse for new workloads. We also worked on the automated deployment of the framework on top of one or multiple IaaS clouds.

9.2.8. PIA ELCI (2015-2018)

Participant: Anne-Cécile Orgerie.

The PIA ELCI project deals with software environment for computation-intensive applications. It is leaded by BULL. In the context of this project, we collaborate with ROMA and Avalon teams from Lyon: we co-supervise a PhD student (Issam Rais) funded by this project with these teams on multicriteria scheduling for large-scale HPC environments.

9.2.9. CNRS PEPS EcoSmart (2016)

Participant: Anne-Cécile Orgerie.

Smart Grids are connected to telecommunication networks and can thus optimize the production, distribution and consumption of electricity. Virtualized distributed systems (Clouds) are the major players in providing services over the Internet. The success of these on-demand services makes the energy consumption of these systems worrying. This project aims to optimize the energy consumption of these large consumers, namely virtualized distributed Clouds consisting of computing, storage and communication resources. The objective is to exploit the capabilities offered by smart grids to control the consumption of these systems and be able to influence it according to the availability or the nature of the electricity used.

9.2.10. CNRS GDS EcoInfo

Participant: Anne-Cécile Orgerie.

The EcoInfo group deals with reducing environmental and societal

impacts of Information and Communications Technologies from hardware to software aspects. This group aims at providing critical studies, lifecycle analyses and best practices in order to improve the energy efficiency of printers, servers, data centers, and any ICT equipment in use in public research organizations.

9.3. European Initiatives

9.3.1. FP7 & H2020 Projects

9.3.1.1. PaaSage

Title: PaaSage: Model Based Cloud Platform Upperware

Programm: FP7

Duration: October 2012 - September 2016

Coordinator: ERCIM

Partners:

- Akademia Gorniczo-Hutnicza Im. Stanislawa Staszica W Krakowie (Poland)
- Automotive Simulation Center Stuttgart Ev (Germany)
- Be.Wan Sprl (Belgium)
- Centred'Excellence en Technologies de l'Information et de la Communication (Belgium)
- Geie Ercim (France)
- Evry As (Norway)
- Flexiant (United Kingdom)
- Foundation for Research and Technology Hellas (Greece)
- Gesellschaft Fur Wissenschaftliche Datenverarbeitung Mbh Gottingen (Germany)
- Ibsac - Intelligent Business Solutions Ltd (Cyprus)
- Inria (France)
- Lufthansa Systems (Germany)
- Stiftelsen Sintef (Norway)
- Science and Technology Facilities Council (United Kingdom)
- University of Cyprus (Cyprus)
- Universitaet Stuttgart (Germany)
- Universitaet Ulm (Germany)
- Universitetet I Oslo (Norway)

Inria contact: Christine Morin

PaaSage (2012-2016) (http://www.paasage.eu) is an FP7 collaborative project that develops an open-source cloud platform, with an accompanying methodology and language, which enables developers to access cloud services in a technology-neutral approach while guiding them to configure their applications for best performance. PaaSage facilitates application deployment on multiple clouds while enhancing the flexibility, adaptivity and scalability of applications. Myriads develops the Adapter subsystem that supports dynamic, cross-cloud application adaptation. In 2016, we improved the Adapter implementation and evaluated its use within the business scenarios of PaaSage partners.

9.3.1.2. Fed4Fire

Title: Federation for FIRE

Programm: FP7

Duration: October 2012 - September 2016

Coordinator: Interdisciplinary institute for broadband technology (iMinds, Belgium)

Partners:

University of Southampton (It Innovation, United Kingdom)

Universite Pierre et Marie Curie - paris6 (UPMC, France)

Fraunhofer-Gesellschaft zur Foerderung der Angewandten Forschung e.v (Fraunhofer, Germany)

Technische Universitat Berlin (TUB, Germany)

The University of Edinburgh (UEDIN, United Kingdom)

National Ict Australia Limited (NICTA, Australia)

Atos Spain SA (Atos, Spain)

Panepistimio Thessalias (University of Thessaly) (UTH, Greece)

National Technical University of Athens (NTUA, Greece)

University of Bristol (UNIVBRIS, United Kingdom)

Fundacio Privada i2cat, Internet I Innovacio Digital a Catalunya (i2cat, Spain)

Eurescom-European Institute for Research and Strategic Studies in Telecommunications (EUR, Gmbh Germany)

Delivery of Advanced Network Technology to Europe limited (DANTE limited, United Kingdom)

Universidad de Cantabria (UC, Spain)

National Information Society agency (NIA, Korea (republic of))

Inria contact: Walid Dabbous

Fed4FIRE is an FP7 Integrated Project project running between October 2012 and September 2016 (http://www.fed4fire.eu), extended to December 2016. In Fed4FIRE, we investigate the means by which our experimental platforms (BonFIRE, and in a secondary way Grid'5000) could be made interoperable with a wider eco-system of experimental platforms in Europe and beyond. 2016 is the sustainability year of Fed4FIRE, and as usage from experimenters is not bringing any revenue, we closed the BonFIRE platform as it was become unmaintainable without significant effort.

9.3.2. Collaborations in European Programs, Except FP7 & H2020

9.3.2.1. NESUS

Participant: Anne-Cécile Orgerie.

Program: ICT COST

Project acronym: NESUS

Project title: Network for Sustainable Ultrascale Computing (ICT COST Action IC1305)

Duration: 2014 - 2018

Coordinator: Prof. Jesus Carretero, University Carlos III of Madrid, Spain, http://www.nesus.eu

Other partners: 33 COST countries and 11 non-COST countries

Abstract: Ultrascale systems are envisioned as large-scale complex systems joining parallel and distributed computing systems that will be two to three orders of magnitude larger that today's systems. The EU is already funding large scale computing systems research, but it is not coordinated across researchers, leading to duplications and inefficiencies. The goal of the NESUS Action is to establish an open European research network targeting sustainable solutions for ultrascale computing aiming at cross fertilization among HPC, large scale distributed systems, and big data management. The network will contribute to gluing disparate researchers working across different areas and provide a meeting ground for researchers in these separate areas to exchange ideas, to identify synergies, and to pursue common activities in research topics such as sustainable software solutions (applications and system software stack), data management, energy efficiency, and resilience. Some of the most active research groups of the world in this area are members of this proposal. This Action will increase the value of these groups at the European-level by reducing duplication of efforts and providing a more holistic view to all researchers, it will promote the leadership of Europe, and it will increase their impact on science, economy, and society. Anne-Cécile Orgerie is co-responsible of the focus group on metrics, monitoring, instrumentation and profiling in the Working Group 5 on Energy Efficiency.

9.3.3. Collaborations with Major European Organizations

Partner 1: EPFL, Network architecture lab (Switzerland)

We collaborate with Katerina Argyraki's research group on the integration of networking and cloud computing technologies in order to support placement constraints between cloud resources.

Partner 2: VU University Amsterdam, dept. of Computer Science (the Netherlands)

We collaborate with Thilo Kielmann's research group at VU University Amsterdam on research and development around the ConPaaS system.

Partner 3: University of Neuchâtel, dept. of Computer Science (Switzerland)

We collaborate with Pascal Felber's research group on energy efficiency in Clouds and in particular on the design of energy cost models for virtual machines.

9.4. International Initiatives

9.4.1. Inria International Labs

9.4.1.1. DALHIS

Title: Data Analysis on Large Heterogeneous Infrastructures for Science

International Partner (Institution - Laboratory - Researcher):

> Lawrence Berkeley National Laboratory (United States) - Data Science and Technology department - Deb Agarwal

Start year: 2016

See also: https://project.inria.fr/dalhis/

Data produced by scientific instruments (large facilities like telescopes or field data), large-scale experiments, and high-fidelity simulations are increasing in magnitude and complexity. Existing data analysis methods, tools and infrastructure are often difficult to use and unable to provide the complete data management, collaboration, and curation environment needed to manage these complex, dynamic, and large-scale data analysis environments. The goal of the Inria-LBL DALHIS associate team involving the Myriads (PI) and Avalon Inria project-teams and the Data Science and Technology (DST) department at Lawrence Berkeley National Laboratory (LBL) is to create a collaborative distributed software ecosystem to manage data lifecycle and enable data analytics on distributed data sets and resources. Specifically, our goal is to build a dynamic software stack that is user-friendly, scalable, energy-efficient and fault tolerant. Our research will determine appropriate execution environments that allow users to seamlessly execute their end-to-end dynamic data analysis workflows in various resource environments and scales while meeting energy-efficiency, performance and fault tolerance goals. We will engage in deep partnerships with scientific teams (Fluxnet in environmental science and SNFactory and LSST experiences in cosmology) and use a mix of user research with system software R&D to address specific challenges that these communities face. Our experience will in turn inform future research directions.

9.4.2. Inria International Partners

9.4.2.1. Informal International Partners

Partner: Rutgers University, dept. of Computer Science (New Jersey, United States)

We collaborate with Manish Parashar's research group on energy efficiency in edge Clouds and in particular on the design of energy cost models for such environments.

Partner: Northeastern University, dept. of Computer Science (Massachusetts, United States)

We collaborate with Gene Cooperman 's research group on virtualization technologies for the study of large-scale distributed systems.

Partner: University of Guadalajara (Mexico)

We collaborate with the team of Prof. H√©ctor Duran-Limon on application and resource management in the cloud. In 2016, we produced a joint journal publication [14]. Nikos Parlavantzas is co-advising a PhD student enrolled in the University of Guadalajara (Carlos Ruiz Diaz).

Partner: Tlemcen University (Algeria)

We collaborate with Djawida Dib on energy-efficient fault-tolerant resource and application management in containerized clouds. Christine Morin will co-advise a PhD student enrolled in the University of Tlemcen (Yasmina Bouizem) from December 2016.

9.5. International Research Visitors

9.5.1. Visits of International Scientists

Deb Agarwal, senior scientist at Lawrence Berkeley National Laboratory, who has been awarded an Inria International Chair for the 2015-2019 period, visited Myriads team during two months from May 1st to June 30th, 2016.

Christian Heinrich, PhD student in the Polaris team at Inria Grenoble, spent one month in October in the Myriads team to improve how large-scala distributed systems are declared in the SimGrid framework.

Professor Gene Cooperman, Northeastern University, Boston, USA, visited the Myriads team for one week in June to revive our collaboration on the virtualization of large-scale distributed systems.

Professor Peter Hubwieser, Technic University of Munchen, Germany, visited the Myriads team for two weeks in November to start a collaboration on the didactics of computer science with Martin Quinson.

Carlos Ruiz Diaz, PhD student in the University of Guadalajara, Mexico, spent 6 months in the Myriads team (from September 2015 to February 2016) to advance his work on adapting cloud configurations.

9.5.1.1. Internships

Benjamin Trubert

Date: May-August 2016

Institution: University of Rennes 1

Supervisor: Guillaume Pierre

Kartik Sathyanarayanan

Date: May-July 2016

Institution: Birla Institute of Technology & Science, Pilani (India)

Supervisor: Christine Morin

9.5.2. Visits to International Teams

9.5.2.1. Research Stays Abroad

- Anna Giannakou did a 3-month research internship in the Data Science and Technology department of the Lawrence Berkeley National Laboratory from August to October 2016. She worked with Sean Peisert, staff scientist, on building a workflow for anomaly Detection in HPC environments using statistical data.
- Yunbo Li did a 2 month research internship in the Computer Science department of Rutgers University from August to September 2016. He worked with Prof. Manish Parashar on building an energy cost model for edge cloud applications.

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events Organisation

10.1.1.1. Member of the Organizing Committees

- Anne-Cécile Orgerie was the main organizer of the research school E3-RSD on energy-efficiency in networks and distributed systems held in Dinard in June 2016 (http://e3rsd.irisa.fr).
- Louis Rilling was the main organizer of the SEC2 2016 workshop on security in Clouds held in Lorient in July 2016 in conjunction with the Compas 2016 conference (https://sec2-2016.inria.fr/).
- Guillaume Pierre was co-workshop/tutorials chair of the ACM/IFIP/USENIX Middleware 2016 conference.
- Martin Quinson was the main organizer of the research school SUD gathering the SimGrid user community held in Fréjus in January.
- Nikos Parlavantzas was a member of the organizing committee for MMS 2016 (INSA Science Day) in April 2016

10.1.2. Scientific Events Selection

10.1.2.1. Chair of Conference Program Committees

- Guillaume Pierre was the program chair of the Systems track of the Compas 2016 national conference.
- Guillaume Pierre was a program co-chair of CloudDP 2016 (6th International Workshop on Cloud Data and Platforms)

10.1.2.2. Member of the Conference Program Committees

- Christine Morin was a member of the program committee of the ACM/IEEE, SC'16, IEEE ICDCS 2016, ACM/IEEE CC-Grid 2016, IEEE CloudCom 2016, IEEE NFV-SDN 2016, Euromicro PDP 2016, ICCCN 2016 conferences and Resilience'16, ScienceCloud'16, CrossCloud'16 workshops.
- Anne-Cécile Orgerie was a member of the program committee of the PASA 2016 workshop and the Compas 2016 conference.
- Guillaume Pierre was a member of the program committees of IEEE IC2E 2016, ICPP 2016, DADS track of ACM SAC 2016, CloudCom 2016, and IEEE ScalCom 2016.
- Martin Quinson was a member of the program committees of SimulTech 2016.
- Nikos Parlavantzas was a member of the program committees of VHPC 16, and CLOSER 2016.
- Cédric Tedeschi was a member of the program committee of ICWS 2016, ICCS 2016, CloSer 2016 and Compas 2016 conferences. He was a member of the program committee of the ERROR 2016 workshop.

10.1.2.3. Reviewer

- Jean-Louis Pazat was a reviewer for the Compas Conference.
- Martin Quinson was a reviewer for the Compas Conference.

10.1.3. Journal

10.1.3.1. Member of the Editorial Boards

Christine Morin is associate editor in the IEEE's Transactions on Parallel and Distributed Systems' Editorial Board.

10.1.3.2. Reviewer - Reviewing Activities

- Cédric Tedeschi was a reviewer for the TPDS journal.

10.1.4. Invited Talks

- "Developing an Understanding of Ecosystems and Energy-Water Interactions Through Data", keynote talk, Inria@Silicon Valley BIS workshop, Paris, June 8th, 2016.
- "Infection dynamics in a metapopulation of cattle herds: models and their execution in large-scale computing infrastructure"; Christine Morin (together with Pauline Ezanno), talk for Rencontres INRA-Inria, Pont Royal en Provence, October 3rd, 2016.
- "Le coût énergétique du stockage dans le Cloud", Anne-Cécile Orgerie, talk for the scientific days of the University of Nantes, Nantes, June 10, 2016.
- "Green Computing and Sustainability", Anne-Cécile Orgerie, Keynote at the Annual Workshop of the French section of URSI (International Union of Radio Science), Rennes, France, March 2016.
- "L'axe Green IT à l'IRISA", Anne-Cécile Orgerie, talk for the HCERES evaluation of IRISA, Rennes, January 21, 2016.
- "Real-time stream data processing for hyper-interactive applications in fog computing platforms", Guillaume Pierre, invited presentation at the US-EU Invited Workshop on the Next Generation Internet of Things, Los Angeles (USA), March 31st 2016.
- "Toward hyper-interactive applications in fog computing platforms", Guillaume Pierre, invited presentation at the 9th Cloud control workshop, Friiberghs Herrgård (Sweden), 27-29 June 2016.
- "Experimental methodologies for large-scale distributed systems", Martin Quinson, research school E3-RSD on energy-efficiency in networks and distributed systems held in Dinard in June 2016.

10.1.5. Leadership within the Scientific Community

- Christine Morin was invited to participate in the CRA-W career mentoring workshop held in Washington DC on November 19-20, 2016. Together with AJ Brush she organized a panel on "Growing Research/Surviving First 2 Years" and together with Lori Pollock she organized a panel on "Networking and Finding Advocates". She also did 1-1 mentoring for junior computer scientists (postdocs, assistant-professors, junior scientists and junior R&D engineers).

- Anne-Cécile Orgerie is co-responsible for the Green axis of the CNRS GDR RSD (Network and Distributed Systems working group).
- Anne-Cécile Orgerie is secretary of the ASF: the French chapter of ACM SIGOPS.
- Cédric Tedeschi is a member of the steering committee of the Compas conference.

10.1.6. Scientific Expertise

Yvon Jégou is a member of the **Comité de Sélection et de Validation** (CSV) of the *Images & Réseaux* cluster (until June 2016).

Christine Morin was a member of the HCERES evaluation committee of the LORIA (visit in Nancy December 13-15, 2016). She was a remote reviewer for the Call 2015 of CHIST-ERA on the topic User-Centric Security, Privacy and Trust in the Internet of Things. She was a member of the Jury for the Inria - Académie des Sciences prizes 2016.

Jean-Louis Pazat is the coordinator of experts in Information Technology for the evaluation of international bilateral collaborations at the ministry of research and education.

Martin Quinson is a member of the scientific committee of the Blaise Pascal foundation for the scientific dissemination of Maths and Computer Science, and also of the 2017 forum “Vivid Maths” (mathématiques vivantes) organized by the CFEM (French commission for the teaching of maths). He acted as an expert for a project of the UNIT foundation, the technical and engineering numerical university.

Nikos Parlavantzas acted as an expert reviewer for ANR and ANRT CIFRE projects.

10.1.7. Research Administration

- Yvon Jégou is a member of the Grid 5000 executive committee (until June 2016).
- Christine Morin is a member of the board of the Project-Team Committee of Inria Rennes Bretagne Atlantique.
- Christine Morin was the scientific coordinator for the Inria evaluation seminar of the "distributed system and middleware" theme, held in Rungis in October 2016.
- Christine Morin has been a member of the University of Rennes 1 board of directors and of the International Affairs Commission and its board since March 2016.
- Anne-Cécile Orgerie is officer (chargée de mission) for the IRISA cross-cutting axis on Green IT.
- Jean-Louis Pazat was the leader of the “Large Scale Systems” department of IRISA until October 2016, after which this responsibility was taken over by Martin Quinson.
- Jean-Louis Pazat was the leader of the IRISA-INSA computing lab until October 2016

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

Christine Morin is responsible for the Internet of Services: Programming Models & Things and Clouds (ISI) teaching unit of the Master in research in Computer Science of the University of Rennes 1 and of the EIT Digital Master School at the University of Rennes 1.

Christine Morin:

- Master 1: FreeRoom project (spanning the two semesters) co-supervised with Cédric Tedeschi and Paul Couderc, University of Rennes 1, France.
- Master 2: 4.5 hours ETD, Energy management in Large Scale Distributed Systems, Centrale-Supelec, Rennes, France.
- Master 2: Internet of Services: Programming Models & Things and Clouds (ISI), 12 hours ETD, EIT Digital Master School, University of Rennes 1, France.

Anne-Cécile Orgerie:

- Licence 3: ARCSYS2: Architecture and System - 24 hours of lecture at ENS Rennes
- Master 1: Cloud & Big Data - 12 hours of lecture and 12 hours of practical sessions at ENS Rennes
- Master 2: Green ICT - 5 hours of invited lecture at Telecom SudParis
- PhD students: Energy issues in HPC and Clouds - 1.5 jours of invited lecture at the research school of the ETN BigStorage

Guillaume Pierre (at the University of Rennes 1):

- License 3: Systèmes (25 hours ETD)
- License 3: Organisation et utilisation des systèmes d'exploitation 2 (67 hours ETD)
- Master 2: Techniques de développement logiciel dans le Cloud (39 hours ETD)
- Master 1: Service Technologies (24 hours ETD)
- Master 2: Approche algorithmique des applications et systèmes répartis (32 hours ETD)
- DU D√©veloppeur-Concepteur Logiciel: Systèmes (42 hours ETD)

Martin Quinson (at ENS Rennes):

- Licence 3: Programming and Software Engineering (30 hours ETD); ARCSYS2 - architecture et syst√®mes 2 (60 hours ETD); Pedagogy (15 hours ETD).
- Agregation Science Industrielle: Programming and Software Engineering (20 hours ETD); Operating Systems and C programming (20 hours ETD); Networking (20 hours ETD).
- Master 2: Pedagogy and Scientific Mediation for Computer Science (30 hours EDT)

Jean-Louis Pazat (at INSA Rennes):

- Licence 3: Parallel and Multicore Programming (36 hours ETD)
- Master 1: Parallel and Distributed Programming (36 hours ETD)
- Master 1: Scientific Parallel Programming (36 hours ETD)
- Licence 3: Networks and SOA (20 hours ETD)

Nikos Parlavantzas (at INSA Rennes):

- Master 1: Performance Evaluation (32 hours ETD)
- Master 1: Operating Systems (36 hours ETD)
- Master 1: Big Data and Applications (15 hours ETD)
- Master 1: Networks and SOA (20 hours ETD)
- Master 1: Advanced Operating Systems (20 hours ETD)
- Master 1: Parallel programming (12 hours ETD)
- Master 1: Software Development Project (30 hours ETD)
- Licence 3: Networking (16 hours ETD)
- Master 2: Component-based Software Engineering (16 hours ETD)

Cédric Tedeschi (220 hours ETD at Univ. Rennes 1):

- Master 1: Cooperation and concurrency in Systems and Networks
- Master 1: Projects (in charge of the *UE*)
- Master 1: Parallel programming
- Master 2: Internet of Services and Infrastructures

10.2.2. Supervision

PhD in progress: Arif Ahmed, "Scalable Decentralized Edge Cloud Infrastructures", started in October 2016, Guillaume Pierre.

PhD in progress: Ismael Cuadrado Cordero, Energy-efficient and network-aware resource allocation in Cloud infrastructures, started in October 2013, Christine Morin, Anne-Cécile Orgerie.

PhD in progress: Anna Giannakou, Self-adaptable security monitoring in Clouds, started in March 2014, Louis Rilling, Christine Morin, Jean-Louis Pazat.

PhD in progress: Baptiste Goupille-Lescar, Designing agile, distributed cyber-physical systems with advanced collaboration capabilities, started in January 2016, Michel Barreteau (Thales) Eric Lenormand (Thales), Christine Morin, Nikos Parlavantzas.

PhD in progress: David Guyon, Supporting energy-awareness for cloud users, started in September 2015, Anne-Cécile Orgerie, Christine Morin.

PhD in progress: Sabbir Hasan, SLA Driven Cloud autoscaling for optimizing energy footprint, started in December 2013, Thomas Ledoux (Ascola), Jean-Louis Pazat.

PhD: Ancuta Iordache, "Performance-cost tradeoffs in heterogeneous clouds", University of Rennes 1, successfully defended on September 9th 2016. Guillaume Pierre.

PhD in progress: Yunbo Li, Resource allocation in a Cloud partially powered by renewable energy sources, started in October 2013, Anne-Cécile Orgerie, Jean-Marc Menaud (Ascola).

PhD in progress: Pernelle Mensah, Security policy adaptation driven by risk evaluation in modern communication infrastructures, started in December 2015, Samuel Dubus (Alcatel-Lucent), Christine Morin, Guillaume Piolle (Cidre), Eric Totel (Cidre).

PhD in progress: Édouard Outin, A multi-objective adaptation system for the management of a Distributed Cloud, started in October 2013, Olivier Barais (Triskell), Yvon Jégou, Jean-Louis Pazat.

PhD in progress: The Anh Pham, Dynamic Formal Verification of High Performance Runtimes and Applications, started in November 2016, Martin Quinson, Thierry Jéron.

PhD in progress: Issam Rais, Multi criteria scheduling for large scale High Performance Computing environments, started in October 2015, Anne-Cécile Orgerie, Anne Benoit (ROMA), Laurent Lefèvre (Avalon).

PhD in progress: Bruno Stevant, Resource allocation strategies for service distribution at the Internet edge to optimize end-to-end latency, started in December 2014, Jean-Louis Pazat.

PhD in progress: Genc Tato, Locality-aware Lazy Overlay Networks for WANS, started in December 2015, Marin Bertier, Cédric Tedeschi, Christine Morin.

PhD in progress: Amir Teshome, Definition and enforcement of Service-Level Agreements for Cloud security monitoring, started in October 2015, Louis Rilling, Christine Morin.

10.2.3. Juries

- Christine Morin was a reviewer in the PhD committee of Maxime Lorrillere, university Pierre et Marie Curie, February 4th, 2016.
- Christine Morin was a reviewer in the PhD committee of Richard Relaza, university of Toulouse, February 12th, 2016.
- Christine Morin was a member in the PhD committee of Anca Iordache, university of Rennes 1, September 9th, 2016.
- Christine Morin chaired the PhD committee of Ji Liu, university of Montpellier, November 3rd, 2016.
- Christine Morin was a member in the PhD committee of Daniel Balouek-Thomert, ENS Lyon, December 5th, 2016.

- Christine Morin chaired the PhD committee of Violaine Villebonnet, ENS Lyon, December 6th, 2016.
- Christine Morin was a reviewer in the PhD committee of Vincente Kherbache, university of Nice, December 7th, 2016.
- Anne-Cécile Orgerie was a member in the PhD committee of Frédéric Dumont, Ecole des Mines de Nantes, 21st of September 2016.
- Anne-Cécile Orgerie was a member in the PhD committee of Maxime Colmant, University of Lille, 24th of November 2016.
- Jean-Louis Pazat chaired the PhD committee of Pierre Obame Meye, University of Rennes 1, December 1st, 2016.
- Jean-Louis Pazat chaired the PhD committee of Fran√ßois Lehericey, INSA Rennes, September 20, 2016.
- Guillaume Pierre was a reviewer in the PhD committee of Luc André, University of Lorraine, May 13th 2016.
- Guillaume Pierre was a reviewer in the HDR committee of Christoph Neumann, University of Rennes 1, November 23rd 2016.
- Guillaume Pierre was a reviewer in the PhD committee of Vincent Kherbache, university of Nice, December 7th, 2016.
- Louis Rilling was a member in the PhD committee of Allan Blanchard, University of Orléans, 6th of December 2016.
- Martin Quinson was a reviewer in the PhD committee of Dimitri Pertin, University of Nantes, April 22th, 2016.
- Martin Quinson was a member in the PhD committee of David Beniamine, University of Grenoble, December 5th, 2016.
- Cédric Tedeschi was a member of the PhD committee of Sajith Kalathingal, University of Rennes 1, December 13th, 2016.

10.3. Popularization

- Martin Quinson is a scientific expert in a teaching manual of Computer Science for primary schools, authored within the "La main à la pâte" popularization network. This manual was released this summer.
- Martin Quinson co-founded a working group of the Société Informatique de France toward the creation of new unplugged activities to teach the informatics to pupils. This group met twice in Lyon this year, in April and November.
- Martin Quinson participated and co-organized to several workshops and open events where high-school students were invited to discover Computer Science (March at ENS Rennes, November on the Beaulieu campus). He was a scientific expert to the MathC2+ event at ENS-Rennes in June.
- Martin Quinson participated to the full day meeting “Splash Education” targeting at determining the fundamental programming concepts that should be taught to every pupils, co-located with the SPLASH conference http://2016.splashcon.org/track/splash-2016-splash-e. He organized the 3-days event on instrumented teaching of computer science, co-located with the Orphee-RDV event https://apprentissageinstrumentdelinformatique.wordpress.com/.
- Martin Quinson continued the development of the PLM web platform, which is an exerciser to teach programming to beginners. He also submitted several project applications to pursue this work in the future. Unfortunately, none of these applications have been accepted so far. Prof Peter Hubwieser (Technical University of Munchen, chair of didactics of Computer Science) visited us for two weeks in November. Developing the PLM and exploiting the data already gathered were central elements of this work meeting. A joint publication is currently prepared, targeting the ItiCSE’17 conference.

11. Bibliography

Major publications by the team in recent years

[1] S. COSTACHE, S. KORTAS, C. MORIN, N. PARLAVANTZAS.*Market-based Autonomous Resource and Application Management in Private Clouds*, in "Journal of Parallel and Distributed Computing", February 2017, vol. 100, p. 85-102 [*DOI :* 10.1016/J.JPDC.2016.10.003], https://hal.archives-ouvertes.fr/hal-01378536.

[2] E. FELLER, L. RAMAKRISHNAN, C. MORIN.*Performance and energy efficiency of big data applications in cloud environments: A Hadoop case study*, in "Journal of Parallel and Distributed Computing", 2015, vol. 79-80, p. 80-89 [*DOI :* 10.1016/J.JPDC.2015.01.001], https://hal.archives-ouvertes.fr/hal-01271141.

[3] E. FELLER, L. RILLING, C. MORIN.*Energy-Aware Ant Colony Based Workload Placement in Clouds*, in "The 12th IEEE/ACM International Conference on Grid Computing (GRID-2011)", Lyon, France, September 2011, https://hal.inria.fr/inria-00626042.

[4] A. IORDACHE, C. MORIN, N. PARLAVANTZAS, P. RITEAU.*Resilin: Elastic MapReduce over Multiple Clouds*, Inria, October 2012, n^{o} RR-8081, https://hal.inria.fr/hal-00737030.

[5] A. IORDACHE, G. PIERRE, P. SANDERS, J. G. D. F. COUTINHO, M. STILLWELL.*High Performance in the Cloud with FPGA Groups*, in "9th IEEE/ACM International Conference on Utility and Cloud Computing (UCC 2016)", Shanghai, China, December 2016, https://hal.inria.fr/hal-01356998.

[6] A.-C. ORGERIE, M. DIAS DE ASUNCAO, L. LEFÈVRE.*A Survey on Techniques for Improving the Energy Efficiency of Large Scale Distributed Systems*, in "ACM Computing Surveys", December 2014, vol. 46, n^{o} 4, To appear, https://hal.inria.fr/hal-00767582.

[7] J. ROJAS BALDERRAMA, M. SIMONIN, C. TEDESCHI.*GinFlow: A Decentralised Adaptive Workflow Execution Manager*, in "30th IEEE International Parallel & Distributed Processing Symposium", Chicago, United States, Proceedings of the 30th IEEE International Parallel & Distributed Processing Symposium, May 2016, https://hal.archives-ouvertes.fr/hal-01292135.

Publications of the year

Doctoral Dissertations and Habilitation Theses

[8] A. IORDACHE.*Performance-cost trade-offs in heterogeneous clouds*, Université Rennes 1, September 2016, https://tel.archives-ouvertes.fr/tel-01419975.

Articles in International Peer-Reviewed Journal

[9] N. BELDICEANU, B. DUMAS FERIS, P. GRAVEY, S. HASAN, C. JARD, T. LEDOUX, Y. LI, D. LIME, G. MADI-WAMBA, J.-M. MENAUD, P. MOREL, M. MORVAN, M.-L. MOULINARD, A.-C. ORGERIE, J.-L. PAZAT, O. H. ROUX, A. SHARAIHA.*Towards energy-proportional Clouds partially powered by renewable energy*, in "Computing", January 2017, vol. 99, n^{o} 1, 20 [*DOI :* 10.1007/s00607-016-0503-z], https://hal.inria.fr/hal-01340318.

[10] M. BERTIER, M. PERRIN, C. TEDESCHI.*On the Complexity of Concurrent Multiset Rewriting*, in "International Journal of Foundations of Computer Science", 2016, vol. 27, n° 1 [*DOI :* 10.1142/S0129054116500052], https://hal.inria.fr/hal-01326849.

[11] E. CARON, A. K. DATTA, C. TEDESCHI, F. PETIT.*Self-Stabilizing Prefix Tree Based Overlay Networks*, in "International Journal of Foundations of Computer Science", 2016, vol. 27, n° 5, p. 607–630 [*DOI :* 10.1142/S0129054116500192], http://hal.upmc.fr/hal-01347457.

[12] S. COSTACHE, S. KORTAS, C. MORIN, N. PARLAVANTZAS.*Market-based Autonomous Resource and Application Management in Private Clouds*, in "Journal of Parallel and Distributed Computing", February 2017, vol. 100, p. 85-102 [*DOI :* 10.1016/J.JPDC.2016.10.003], https://hal.archives-ouvertes.fr/hal-01378536.

[13] M. DORIER, O. YILDIZ, S. IBRAHIM, A.-C. ORGERIE, G. ANTONIU.*On the energy footprint of I/O management in Exascale HPC systems*, in "Future Generation Computer Systems", March 2016, vol. 62, p. 17–28 [*DOI :* 10.1016/J.FUTURE.2016.03.002], https://hal.inria.fr/hal-01330735.

[14] H. A. DURAN-LIMON, J. FLORES-CONTRERAS, N. PARLAVANTZAS, M. ZHAO, A. MEULENERT-PEÑA.*Efficient execution of the WRF model and other HPC applications in the cloud*, in "Earth Science Informatics", March 2016 [*DOI :* 10.1007/S12145-016-0253-7], https://hal.archives-ouvertes.fr/hal-01288718.

[15] H. FERNANDEZ, C. TEDESCHI, T. PRIOL.*A Chemistry-Inspired Workflow Management System for Decentralizing Workflow Execution*, in "IEEE Transactions on Services Computing", 2016, vol. 9, n° 2 [*DOI :* 10.1109/TSC.2013.27], https://hal.inria.fr/hal-01326878.

[16] A. LAGE FREITAS, N. PARLAVANTZAS, J.-L. PAZAT.*Cloud resource management driven by profit augmentation*, in "Concurrency and Computation: Practice and Experience", June 2016 [*DOI :* 10.1002/CPE.3899], https://hal.archives-ouvertes.fr/hal-01337996.

Invited Conferences

[17] A.-C. ORGERIE.*Green Computing and Sustainability*, in "Energie et radiosciences - Journées scientifiques URSI France", Rennes, France, March 2016, https://hal.inria.fr/hal-01356921.

International Conferences with Proceedings

[18] N. CHERIERE, P. DONAT-BOUILLUD, S. IBRAHIM, M. SIMONIN.*On the Usability of Shortest Remaining Time First Policy in Shared Hadoop Clusters*, in "SAC 2016-The 31st ACM/SIGAPP Symposium on Applied Computing", Pisa, Italy, April 2016, https://hal.inria.fr/hal-01239341.

[19] I. CUADRADO CORDERO, F. CUADRADO, C. PHILLIPS, A.-C. ORGERIE, C. MORIN.*Microcities: a Platform based on Microclouds for Neighborhood Services*, in "International Conference on Algorithms and Architectures for Parallel Processing (ICA3PP)", Granada, Spain, December 2016, https://hal.inria.fr/hal-01362532.

[20] I. CUADRADO-CORDERO, A.-C. ORGERIE, C. MORIN.*Incentives for Mobile Cloud Environments through P2P Auctions*, in "CloudNet: IEEE International Conference on Cloud Networking", Pisa, Italy, October 2016, https://hal.inria.fr/hal-01355685.

[21] A. GIANNAKOU, L. RILLING, J.-L. PAZAT, C. MORIN.*AL-SAFE: A Secure Self-Adaptable Application-Level Firewall for IaaS Clouds*, in "CloudCom2016-8th IEEE International Conference on Cloud Computing Technology and Science", luxembourg, Luxembourg, December 2016, https://hal.inria.fr/hal-01363540.

[22] D. GUYON, A.-C. ORGERIE, C. MORIN, D. AGARWAL.*How Much Energy can Green HPC Cloud Users Save?*, in "PDP 2017 - 25th Euromicro International Conference on Parallel, Distributed, and Network-Based Processing", Saint Petersbourg, Russia, March 2017, https://hal.inria.fr/hal-01439874.

[23] M. S. HASAN, F. ALVARES DE OLIVEIRA, T. LEDOUX, J.-L. PAZAT.*Enabling Green Energy awareness in Interactive Cloud Application*, in "IEEE International Conference on Cloud Computing Technology and Science 2016", Luxembourg, Luxembourg, December 2016, https://hal.inria.fr/hal-01365230.

[24] A. IORDACHE, G. PIERRE, P. SANDERS, J. G. D. F. COUTINHO, M. STILLWELL.*High Performance in the Cloud with FPGA Groups*, in "9th IEEE/ACM International Conference on Utility and Cloud Computing (UCC 2016)", Shanghai, China, December 2016, https://hal.inria.fr/hal-01356998.

[25] M. KURPICZ, A.-C. ORGERIE, A. SOBE.*How much does a VM cost? Energy-proportional Accounting in VM-based Environments*, in "PDP: Euromicro International Conference on Parallel, Distributed, and Network-Based Processing", Heraklion, Greece, February 2016, 8 [*DOI :* 10.1109/PDP.2016.70], https://hal.inria.fr/hal-01276913.

[26] Y. LI, A.-C. ORGERIE, J.-M. MENAUD.*Balancing the use of batteries and opportunistic scheduling policies for maximizing renewable energy consumption in a Cloud data center*, in "PDP 2017 - 25th Euromicro International Conference on Parallel, Distributed, and Network-Based Processing", St Petersburg, Russia, March 2017, https://hal.inria.fr/hal-01432752.

[27] A. LÈBRE, A. SIMONET, A.-C. ORGERIE.*Deploying Distributed Cloud Infrastructures: Who and at What Cost?*, in "Proceedings of the fifth IEEE International Workshop on Cloud Computing Interclouds, Multiclouds, Federations, and Interoperability", Berlin, Germany, Intercloud 2016, April 2016, 6 [*DOI :* 10.1109/IC2EW.2016.48], https://hal.inria.fr/hal-01404594.

[28] E. OUTIN, J.-E. DARTOIS, O. BARAIS, J.-L. PAZAT.*Seeking for the Optimal Energy Modelisation Accuracy to Allow Efficient Datacenter Optimizations*, in "16th International Symposium on Cluster, Cloud and Grid Computing", Cartagena, Italy, IEEE/ACM 16th International Symposium on Cluster, Cloud and Grid Computing, May 2016 [*DOI :* 10.1109/CCGRID.2016.67], https://hal.archives-ouvertes.fr/hal-01356099.

[29] I. RAIS, L. LEFÈVRE, A. BENOIT, A.-C. ORGERIE.*An Analysis of the Feasibility of Energy Harvesting with Thermoelectric Generators on Petascale and Exascale Systems*, in "Workshop Optimization of Energy Efficient HPC & Distributed Systems (OPTIM 2016) - The 2016 International Conference on High Performance Computing & Simulation (HPCS 2016)", Innsbruck, Austria, Proceedings of the 2016 International Conference on High Performance Computing & Simulation (HPCS 2016), July 2016, https://hal.inria.fr/hal-01348554.

[30] I. RAIS, A.-C. ORGERIE, M. QUINSON.*Impact of Shutdown Techniques for Energy-Efficient Cloud Data Centers*, in "ICA3PP: 16th International Conference on Algorithms and Architectures for Parallel Processing", Granada, Spain, December 2016, p. 203-210, https://hal.inria.fr/hal-01362530.

[31] J. ROJAS BALDERRAMA, M. SIMONIN, C. TEDESCHI.*Chemistry-Inspired Adaptive Stream Processing*, in "16th International Conference on Membrane Computing", Valencia, Spain, August 2016, https://hal.inria.fr/hal-01326884.

[32] J. ROJAS BALDERRAMA, M. SIMONIN, C. TEDESCHI.*GinFlow: A Decentralised Adaptive Workflow Execution Manager*, in "30th IEEE International Parallel & Distributed Processing Symposium", Chicago, United States, Proceedings of the 30th IEEE International Parallel & Distributed Processing Symposium, May 2016, https://hal.archives-ouvertes.fr/hal-01292135.

[33] C. RUIZ, H. A. DURAN-LIMON, N. PARLAVANTZAS.*Towards a software product line-based approach to adapt IaaS cloud configurations*, in "9th International Conference on Utility and Cloud Computing (UCC '16)", Shanghai, China, December 2016 [*DOI :* 10.1145/2996890.3007893], https://hal.archives-ouvertes.fr/hal-01433820.

[34] A. TESHOME, L. RILLING, C. MORIN.*Including Security Monitoring in Cloud Service Level Agreements*, in "Symposium on Reliable Distributed Systems (SRDS), PhD forum", Budapest, Hungary, September 2016, https://hal.inria.fr/hal-01354975.

Conferences without Proceedings

[35] A. GIANNAKOU, L. RILLING, C. MORIN, J.-L. PAZAT.*AL-SAFE: A Secure Self-Adaptable Application-Level Firewall for IaaS Clouds*, in "SEC2 2016 - Second workshop on Security in Clouds", Lorient, France, July 2016, https://hal.inria.fr/hal-01340494.

[36] A. GIANNAKOU, L. RILLING, C. MORIN, J.-L. PAZAT.*How to Secure Application-Level Firewalls in IaaS Clouds*, in "CloudDP, 6th International Workshop on Cloud Data and Platforms", London, United Kingdom, Guillaume Pierre, Rui Oliveira , April 2016, https://hal.inria.fr/hal-01340473.

[37] I. RAIS, A. BENOIT, L. LEFÈVRE, A.-C. ORGERIE.*An Analysis of the Feasibility of Energy Harvesting with Thermoelectric Generators on Petascale and Exascale Systems*, in "Conférence d'informatique en Parallélisme, Architecture et Système (COMPAS 2016)", Lorient, France, Actes de COMPAS, la Conférence d'informatique en Parallélisme, Architecture et Système, July 2016, https://hal.inria.fr/hal-01348555.

[38] L. RILLING, C. MORIN.*INDIC : Supervision de sécurité à la carte pour les utilisateurs de cloud IaaS*, in "Rendez-vous de la Recherche et de l'Enseignement de la Sécurité des systèmes d'information (RESSI 2016)", Toulouse, France, LAAS-CNRS, INSA Toulouse, May 2016, https://hal.inria.fr/hal-01403774.

[39] A. TESHOME, L. RILLING, C. MORIN.*Including Security Monitoring in Cloud SLA*, in "SEC2 2016 - Second workshop on Security in Clouds", Lorient, France, July 2016, https://hal.inria.fr/hal-01354974.

Research Reports

[40] M. CALLAU-ZORI, L. SAMOILA, A.-C. ORGERIE, G. PIERRE.*An experiment-driven energy consumption model for virtual machine management systems*, IRISA ; Université de Rennes 1 ; CNRS, January 2016, n^{o} RR-8844, https://hal.inria.fr/hal-01258766.

[41] R.-A. CHERRUEAU, A. LEBRE, D. PERTIN, A. SIMONET, M. SIMONIN.*ENOS: a HolisticFramework forConducting ScientificEvaluations of OpenStack*, Inria Rennes Bretagne Atlantique ; Nantes, November 2016, n^{o} RT-0485, https://hal.inria.fr/hal-01415522.

[42] I. CUADRADO-CORDERO, F. CUADRADO, C. PHILLIPS, A.-C. ORGERIE, C. MORIN.*Microcities: a Platform based on Microclouds for Neighborhood Services*, inria, February 2016, n° RR-8885, https://hal.inria.fr/hal-01291536.

Other Publications

[43] A. DEGOMME, A. LEGRAND, M. S. MARKOMANOLIS, M. QUINSON, M. S. STILLWELL, F. S. SUTER.*Simulating MPI applications: the SMPI approach*, November 2016, Under review in IEEE TPDS, https://hal.inria.fr/hal-01415484.

References in notes

[44] N. BELDICEANU, B. DUMAS FERIS, P. GRAVEY, S. HASAN, C. JARD, T. LEDOUX, Y. LI, D. LIME, G. MADI-WAMBA, J.-M. MENAUD, P. MOREL, M. MORVAN, M.-L. MOULINARD, A.-C. ORGERIE, J.-L. PAZAT, O. ROUX, A. SHARAIHA.*The EPOC project: Energy Proportional and Opportunistic Computing system*, in "International Conference on Smart Cities and Green ICT Systems (SMARTGREENS)", Lisbonne, Portugal, May 2015, 15397, https://hal.archives-ouvertes.fr/hal-01131602.

[45] A. BSILA, N. FERRY, G. HORN, T. KIRKHAM, M. MALAWSKI, N. PARLAVANTZAS, C. PÉREZ, J. ROUZAUD-CORNABAS, D. ROMERO, A. ROSSINI, A. SOLBERG, H. SONG.*D3.1.1 - Upperware Prototype*, March 2014, https://hal.inria.fr/hal-00991340.

[46] I. CUADRADO CORDERO, A.-C. ORGERIE, C. MORIN.*GRaNADA: A Network-Aware and Energy-Efficient PaaS Cloud Architecture*, in "IEEE International Conference on Green Computing and Communications (GreenCom)", Sydney, Australia, December 2015, https://hal.archives-ouvertes.fr/hal-01205905.

[47] D. GUYON.*Energy-efficient Cloud Elasticity for Data-driven Applications*, Université de Rennes 1, June 2015, 46, https://hal.inria.fr/hal-01212524.

[48] D. GUYON, A.-C. ORGERIE, C. MORIN.*Energy-efficient User-oriented Cloud Elasticity for Data-driven Applications*, in "IEEE International Conference on Green Computing and Communications (GreenCom)", Sydney, Australia, December 2015, https://hal.archives-ouvertes.fr/hal-01205915.

[49] A. IORDACHE, E. BUYUKKAYA, G. PIERRE.*Heterogeneous Resource Selection for Arbitrary HPC Applications in the Cloud*, in "Distributed Applications and Interoperable Systems", Grenoble, France, June 2015 [*DOI :* 10.1007/978-3-319-19129-4_9], https://hal.inria.fr/hal-01159024.

[50] A. IORDACHE.*Performance-Cost Trade-Offs in Heterogeneous Clouds*, Université de Rennes 1, France, September 2016, https://hal.inria.fr/tel-01419975.

[51] Y. LI, A.-C. ORGERIE, J.-M. MENAUD.*Opportunistic Scheduling in Clouds Partially Powered by Green Energy*, in "IEEE International Conference on Green Computing and Communications (GreenCom)", Sydney, Australia, December 2015, https://hal.archives-ouvertes.fr/hal-01205911.

[52] A. LÈBRE, J. PASTOR, M. BERTIER, F. DESPREZ, J. ROUZAUD-CORNABAS, C. TEDESCHI, P. ANEDDA, G. ZANETTI, R. NOU, T. CORTES, E. RIVIÈRE, T. ROPARS.*Beyond The Cloud, How Should Next Generation Utility Computing Infrastructures Be Designed?*, Inria, July 2013, n° RR-8348, https://hal.inria.fr/hal-00854204.

[53] K. RAZAVI, A. ION, G. TATO, K. JEONG, R. FIGUEIREDO, G. PIERRE, T. KIELMANN.*Kangaroo: A Tenant-Centric Software-Defined Cloud Infrastructure*, in "Proceedings of the IEEE International Conference on Cloud Engineering", Tempe, AZ, USA, United States, March 2015, https://hal.inria.fr/hal-01100774.

Project-Team PACAP

Pushing Architecture and Compilation for Application Performance

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

IN PARTNERSHIP WITH:
Université Rennes 1

RESEARCH CENTER
Rennes - Bretagne-Atlantique

THEME
Architecture, Languages and Compilation

Table of contents

Project-Team PACAP

Creation of the Project-Team: 2016 July 01

Keywords:

Computer Science and Digital Science:

1.1. - Architectures
1.1.1. - Multicore
1.1.2. - Hardware accelerators (GPGPU, FPGA, etc.)
1.1.3. - Memory models
1.1.4. - High performance computing
1.1.5. - Exascale
1.1.9. - Fault tolerant systems
1.1.10. - Reconfigurable architectures
1.6. - Green Computing
2.2. - Compilation
2.2.1. - Static analysis
2.2.2. - Memory models
2.2.3. - Run-time systems
2.2.4. - Parallel architectures
2.2.5. - GPGPU, FPGA, etc.
2.2.6. - Adaptive compilation
2.3.1. - Embedded systems
2.3.3. - Real-time systems
4.2. - Correcting codes
4.4. - Security of equipment and software
7.11. - Performance evaluation
7.12. - Computer arithmetic

Other Research Topics and Application Domains:

1. - Life sciences
2. - Health
3. - Environment and planet
4. - Energy
5. - Industry of the future
6. - IT and telecom
7. - Transport and logistics
8. - Smart Cities and Territories
9. - Society and Knowledge

1. Members

Research Scientists

Erven Rohou [Team leader, Inria, Senior Researcher, HDR]

Sylvain Collange [Inria, Researcher]
Pierre Michaud [Inria, Researcher]
André Seznec [Inria, Senior Researcher, HDR]

Faculty Members
Damien Hardy [Univ. Rennes I, Associate Professor]
Isabelle Puaut [Univ. Rennes I, Professor, HDR]

Technical Staff
Nicolas Kiss [Inria, from Apr 2016]
Imane Lasri [Inria, from Mar 2016]
Sébastien Martinez [Univ. Rennes I, from June 2016]
Arthur Perais [Inria]
Emmanuel Riou [Inria, until Nov 2016]

PhD Students
Arif Ali Ana-Pparakkal [Inria]
Rabab Bouziane [Inria]
Nabil Hallou [Inria, until Nov 2016]
Kleovoulos Kalaitzidis [Inria, from Jun 2016]
Sajith Kalathingal [Inria, until Oct 2016]
Andrea Mondelli [Inria]
Viet Anh Nguyen [Univ. Rennes I]
Benjamin Rouxel [Univ. Rennes I]
Aswinkumar Sridharan [Inria]
Arjun Suresh [Inria, until Jul 2016]

Post-Doctoral Fellows
Fernando Endo [Inria]
Biswabandan Panda [Inria]

Visiting Scientists
Stefano Cherubin [Politecnico di Milano, from Sep 2016 until Oct 2016]
Anita Tino [Ryerson University, from Oct 2016]
Rubens Emilio Alves Moreira [Universidade Federal de Minas Gerais, from Feb 2016 to May 2016]

Administrative Assistant
Virginie Desroches [Inria]

2. Overall Objectives

2.1. Overall Objectives

2.1.1. Long-Term Goal

In brief, the long-term goal of the PACAP project-team is about *performance*, that is: how fast programs run. We intend to contribute to the ongoing race for exponentially increasing performance and for performance guarantees.

Traditionally, the term "performance" is understood as "how much time is needed to complete execution". *Latency*-oriented techniques focus on minimizing the average-case execution time (ACET). We are also interested in other definitions of performance. *Throughput*-oriented techniques are interested in how many units of computations can be completed per unit of time. This is more relevant on manycores and GPUs where many computing nodes are available, and latency is less critical. Finally, we also study worst-case execution times (WCET). They are extremely important for critical real-time systems where designers must guarantee that deadlines are met, in any situation.

Given the complexity of current systems, simply assessing their performance has become a non-trivial task which we also plan to tackle.

We occasionally consider other metrics related to performance, such as power efficiency, total energy, overall complexity, and real-time response guarantee. Our ultimate goal is to propose solutions that make computing systems more efficient, taking into account current and envisioned applications, compilers, runtimes, operating systems, and microarchitectures. And since increased performance often comes at the expense of another metric, identifying the related trade-offs is of interest to PACAP.

ALF witnessed the end of the "magically" increasing clock frequency and the introduction of commodity multicore processors. PACAP will likely experience the end of Moore's law [0], and the generalization of commodity heterogeneous manycore processors. This impacts how performance is increased and how it can be guaranteed. It is also a time where exogenous parameters should be promoted to first-class citizens:

1. the existence of faults, whose impact is becoming increasingly important when the photo-lithography feature size decreases;
2. the need for security at all levels of computing systems;
3. *green* computing, or the growing concern of power consumption.

2.1.2. Approach

We strive to address performance in a way as transparent as possible for users. For example, instead of proposing any new language, we consider existing applications (written for example in standard C), and we develop compiler optimizations that immediately benefit programmers; we propose microarchitectural features as opposed to changes in processor instruction sets; we analyze and re-optimize binary programs automatically, without any user intervention.

The perimeter of research directions proposed for the PACAP project-team derive from the intersection of two axes: on the one hand, our high-level research objectives, derived from the overall panorama of computing systems, on the other hand the existing expertise and background of the team members on key technology (see illustration on Figure 1). Note that it does not imply that we will systematically explore all intersecting points of the figure, yet all correspond to a sensible research direction. These lists are neither exhaustive, nor final. Operating systems in particular constitute a promising operating point for several of the issues we plan to tackle. Other aspects will likely emerge during the lifespan of the project-team.

2.1.3. Latency-oriented Computing

Improving the ACET of general purpose systems has been the core "business" of CAPS and ALF for two decades. We plan to pursue this line of research, acting at all levels: compilation, dynamic optimizations, and microarchitecture.

2.1.4. Throughput-Oriented Computing

The goal is to maximize the performance-to-power ratio. We will leverage the execution model of throughput-oriented architectures (such as GPUs) and extend it towards general purpose systems. To address the memory wall issue, we will consider bandwidth saving techniques, such as cache and memory compression.

2.1.5. Real-Time Systems – WCET

Designers of real-time systems must provide an upper bound of the worst-case execution time of the tasks within their systems. By definition this bound must be safe (i.e. greater than any possible execution time). To be useful, WCET estimates have to be as tight as possible. The process of obtaining a WCET bound consists in analyzing a binary executable, modeling the hardware, and then maximizing an objective function that takes into account all possible flows of execution and their respective execution times. Our research will consider the following directions:

[0] Moore's law states that the number of transistors in a circuit doubles (approximately) every two years.

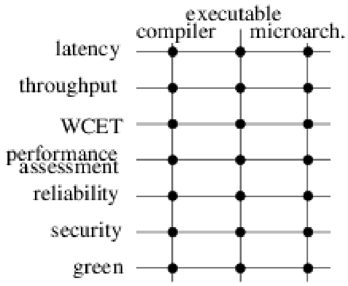


Figure 1. Perimeter of Research Objectives

1. better modeling of hardware to either improve tightness, or handle more complex hardware (e.g. multicores);
2. eliminate unfeasible paths from the analysis;
3. consider probabilistic approaches where WCET estimates are provided with a confidence level.

2.1.6. Performance Assessment

Moore's law drives the complexity of processor micro-architectures, which impacts all other layers: hypervisors, operating systems, compilers and applications follow similar trends. While a small category of experts is able to comprehend (parts of) the behavior of the system, the vast majority of users are only exposed to – and interested in – the bottom line: how fast their applications are actually running. In the presence of virtual machines and cloud computing, multi-programmed workload add yet another degree of non-determinism to the measure of performance. We plan to research how application performance can be characterized and presented to a final user: behavior of the microarchitecture, relevant metrics, possibly visual rendering. Targeting our own community, we also research techniques appropriate for fast and accurate ways to simulate future architectures, including heterogeneous designs, such as latency/throughput platforms.

Once diagnosed, the way bottlenecks are addressed depends on the level of expertise of users. Experts can typically be left with a diagnostic as they probably know better how to fix the issue. Less knowledgeable users must be guided to a better solution. We plan to rely on iterative compilation to generate multiple versions of critical code regions, to be used in various runtime conditions. To avoid the code bloat resulting from multiversioning, we will leverage split-compilation to embed code generation "recipes" to be applied just-in-time, or even at rutime thanks to dynamic binary translation. Finally, we will explore the applicability of auto-tuning, where programmers expose which parameters of their code can be modified to generate alternate versions of the program (for example trading energy consumption for quality of service) and let a global orchestrator make decisions.

2.1.7. Dealing with Faults – Reliability

Semiconductor technology evolution suggests that permanent failure rates will increase dramatically with scaling. While well-known approaches, such as error correcting codes, exist to recover from failures and provide fault-free chips, the exponential growth of the number of faults will make them unaffordable in the future. Consequently, other approaches like fine-grained disabling and reconfiguration of hardware elements (e.g. individual functional units or cache blocks) will become economically necessary. This fine-grained disabling will degrade performance compared to a fault-free execution. This evolution impacts performance (both ACET and WCET). We plan to address this evolution, and propose new techniques, which can be developed at any level. For example, at microarchitecture level, one might consider designing part of a cache in an older technology to guarantee a minimum level of performance; at compile-time, one might generate redundant code for critical sections; at run-time, one can monitor faults and apply corrective measures to the software, or hardware. Solutions involving multiple levels are also very promising.

2.1.8. Dealing with Attacks – Security

Computer systems are under constant attack, from young hackers trying to show their skills, to "professional" criminals stealing credit card information, and even government agencies with virtually unlimited resources. A vast amount of techniques have been proposed in the literature to circumvent attacks. Many of them cause significant slowdowns due to additional checks and countermeasures. Thanks to our expertise in microarchitecture and compilation techniques, we will be able to significantly improve efficiency, robustness and coverage of security mechanism, as well as to partner with field experts to design innovative solutions.

2.1.9. Green Computing

Power consumption has become a major concern of computing systems, at all form factors, ranging from energy-scavenging sensors for IoT, to battery powered embedded systems and laptops, and up to supercomputers operating in the tens of megawatts. Execution time and energy are often related optimization goals. Optimizing for performance under a given power cap, however, introduces new challenges. It also turns out that technologists introduce new solutions (e.g. magnetic RAM) which, in turn, result in new trade-offs and optimization opportunities.

3. Research Program

3.1. Motivation

Our research program is naturally driven by the evolution of our ecosystem. Relevant recent changes can be classified in the following categories: technological constraints, evolving community, and domain constraints. We hereby summarize these evolutions.

3.1.1. Technological constraints

Until recently, binary compatibility guaranteed portability of programs, while increased clock frequency and improved micro-architecture provided increased performance. However, in the last decade, advances in technology and micro-architecture started translating into more parallelism instead. Technology roadmaps even predict the feasibility of thousands of cores on a chip by 2020. Hundreds are already commercially available. Since the vast majority of applications are still sequential, or contain significant sequential sections, such a trend put an end to the automatic performance improvement enjoyed by developers and users. Many research groups consequently focused on parallel architectures and compiling for parallelism.

The focus of ALF – and the DAL ERC – was paradoxically on Amdahl's law: the performance of applications will ultimately be driven by the performance of the sequential part. Despite a number of advances (some of them contributed by ALF), sequential tasks are still a major performance bottleneck. Addressing it is still on the agenda of the proposed PACAP project-team.

In addition, due to power constraints, only part of the billions of transistors of a microprocessor can be operated at any given time (the *dark silicon* paradigm). A sensible approach consists in specializing parts of the silicon area to provide dedicated accelerators (not run simultaneously). This results in diverse and heterogeneous processor cores. Application and compiler designers are now confronted with a moving target, challenging portability and jeopardizing performance.

Finally, we live in a world where billions of sensors, actuators, and computers play a crucial role in our life: flight control, nuclear plant management, defense systems, banking, or health care. These systems must be reliable, despite the fact that they are subject to faults (for example due to aging, charged particle hit, or random noise). Faults will soon become the new *de facto* standard. The evolutions of the semiconductor industry predict an exponential growth of the number of permanent faults [58]. Reliability considerations usually degrade performance. We will propose solutions to mitigate this impact (for example by limiting overheads to critical sections).

Note on technology.
Technology also progresses at a fast pace. We do not propose to pursue any research on technology *per se*. Recently proposed paradigms (quantum computing, non-Si, brain-inspired) have received lots of attention from the research community. We do *not* intend to invest in those paradigms, but we will continue to investigate compilation and architecture for more conventional programming paradigms. Still, several technological shifts may have consequences for us, and we will closely monitor their developments, they include for example non-volatile memory (impacts security, makes writes longer than loads), 3D-stacking (impacts bandwidth), and photonics (impacts latencies and connection network).

3.1.2. Evolving community

The PACAP project-team will tackle performance-related issues, for conventional programming paradigms. In fact, programming complex environments is no longer reserved to experts in compilation and architecture. A large community now develops applications for a wide range of targets, including mobile "apps", cloud, multicore or heterogeneous processors.

This also includes domain scientists (in biology, medicine, but also social sciences) who started relying heavily on computational resources, gathering huge amounts of data, and requiring considerable amount of processing to analyze them. Our research is motivated by the growing discrepancy between on the one hand the complexity of the workloads and the computing systems, and on the other hand the expanding community of developers at large, with limited expertise to optimize and to map efficiently computations to compute nodes.

3.1.3. Domain constraints

Mobile, embedded systems have become ubiquitous. Many of them have real-time constraints. For this class of systems, correctness implies not only producing the correct result, but also doing so within specified deadlines. In the presence of heterogeneous, complex and highly dynamic systems, producing *tight* (i.e. useful) upper bound to the worst-case execution time has become extremely challenging. Our research will aim at improving the tightness as well as enlarging the set of features that can be safely analyzed.

The ever growing dependence of our economy on computing systems also implies that security has become of utmost importance. Many systems are under constant attacks from intruders. Protection has a cost also in terms of performance. We plan to leverage our background to contribute solutions that minimize this impact.

Note on Applications Domains.
As was already the case for ALF, PACAP will work on fundamental technologies for computer science: processor architecture, performance-oriented compilation and guaranteed response time for real-time. The research results may have impacts on any application domain that requires high performance execution (telecommunication, multimedia, biology, health, engineering, environment...), but also on many embedded applications that exhibit other constraints such as power consumption, code size and guaranteed response time.

We strive to extract from active domains the fundamental characteristics that are relevant to our research. For example, *big data* is of interest to PACAP because it relates to the study of hardware/software mechanisms to efficiently transfer huge amounts of data to the computing nodes. Similarly, the *Internet of Things* is of interest because it has implications in terms of ultra low power consumption.

3.2. Research Objectives

Processor micro-architecture and compilation have been at the core of the research carried by the CAPS and ALF project teams for two decades, with undeniable contributions. They will continue to be the foundation of PACAP.

Heterogeneity and diversity of processor architectures now require new techniques to guarantee that the hardware is satisfactorily exploited by the software. We will devise new static compilation techniques (cf. Section 3.2.1), but also build upon iterative [1] and split [2] compilation to continuously adapt software to its environment (Section 3.2.2). Dynamic binary optimization will also play a key role in delivering adapting software and delivering performance.

The end of Moore's law and Dennard's scaling [0] offer an exciting window of opportunity, where performance improvements will no longer derive from additional transistor budget or increased clock frequency, but rather come from breakthroughs in microarchitecture (Section 3.2.3). We will also consider how to reconcile CPU and GPU designs (Section 3.2.4).

Heterogeneity and multicores are also major obstacles to determining tight worst-case execution times of real-time systems (Section 3.2.5), which we plan to tackle.

Finally, we also describe how we plan to address transversal aspects such reliability (Section 3.2.6), power efficiency (Section 3.2.7), and security (Section 3.2.8).

3.2.1. Static Compilation

Static compilation techniques will continue to be relevant to address the characteristics of emerging hardware technologies, such as non-volatile memories, 3D-stacking, or novel communication technologies. These techniques expose new characteristics to the software layers. As an example, non-volatile memories typically have asymmetric read-write latencies (writes are much longer than reads) and different power consumption profiles. PACAP will study the new optimization opportunities and develop tailored compilation techniques for the upcoming compute nodes. New technologies may also be coupled with traditional solutions to offer new trade-offs. We will study how programs can adequately exploit the specific features of the proposed heterogeneous compute nodes.

We propose to build upon iterative compilation [1] to explore how applications perform on different configurations. When possible, Pareto points will be related to application characteristics. The best configuration, however, may actually depend on runtime information, such as input data, dynamic events, or properties that are available only at runtime. Unfortunately a runtime system has little time and means to determine the best configuration. For these reasons, we will also leverage split-compilation [2]: the idea consists in pre-computing alternatives, and embedding in the program enough information to assist and drive a runtime system towards to the best solution.

3.2.2. Software Adaptation

More than ever, software will need to adapt to their environment. In most cases, this environment will remain unknown until runtime. This is already the case when one deploys an application to a cloud, or an "app" to mobile devices. The dilemma is the following: for maximum portability, developers should target the most general device; but for performance they would like to exploit the most recent and advanced hardware features. JIT compilers can handle the situation to some extent, but binary deployment requires dynamic binary rewriting. Our work has shown how SIMD instructions can be upgraded from SSE to AVX [3]. Many more opportunities will appear with diverse and heterogeneous processors, featuring various kinds of accelerators.

On shared hardware, the environment is also defined by other applications competing for the same computational resources. It will become increasingly important to adapt to changing runtime conditions, such as the contention of the cache memories, available bandwidth, or hardware faults. Fortunately, optimizing at runtime is also an opportunity, because this is the first time the program is visible as a whole: executable and libraries (including library versions). Optimizers may also rely on dynamic information, such as actual input data, parameter values, etc. We have already developed a software platform [14] to analyze and optimize programs at runtime, and we started working on automatic dynamic parallelization of sequential code, and dynamic specialization.

We started addressing some of these challenges in ongoing projects such as Nano2017 PSAIC Collaborative research program with STMicroelectronics, as well as within the Inria Project Lab MULTICORE. The starting H2020 FET HPC project ANTAREX will also address these challenges from the energy perspective. We will further leverage our platform and initial results to address other adaptation opportunities. Efficient software adaptation will require expertise from all domains tackled by PACAP, and strong interaction between all team members is expected.

[0] According to Dennard scaling, as transistors get smaller the power density remains constant, and the consumed power remains proportional to the area.

3.2.3. Research directions in uniprocessor microarchitecture

Achieving high single-thread performance remains a major challenge even in the multicore era (Amdahl's law). The members of the PACAP project-team have been conducting research in uniprocessor micro-architecture research for about 20 years covering major topics including caches, instruction front-end, branch prediction, out-of-order core pipeline, branch prediction and value prediction. In particular, in the recent years they have been recognized world leaders in branch prediction [19][9] and in cache prefetching [7] and they have revived the forgotten concept of value prediction [12], [11]. This research was supported by the ERC Advanced grant DAL (2011-2016) and also by Intel. We intend to pursue research on achieving ultimate unicore performance. Below are several non-orthogonal directions that we have identified for mid-term research:

1. management of the memory hierarchy (particularly the hardware prefetching);
2. practical design of very wide issue execution core;
3. speculative execution.

Memory design issues:
Performance of many applications is highly impacted by the memory hierarchy behavior. The interactions between the different components in the memory hierarchy and the out-of-order execution engine have high impact on performance.

The last *Data Prefetching Contest* held with ISCA 2015 has illustrated that achieving high prefetching efficiency is still a challenge for wide-issue superscalar processors, particularly those featuring a very large instruction window. The large instruction window enables an implicit data prefetcher. The interaction between this implicit hardware prefetcher and the explicit hardware prefetcher is still relatively mysterious as illustrated by Pierre Michaud's BO prefetcher (winner of DPC2) [7]. The first objective of the research is to better understand how the implicit prefetching enabled by the large instruction window interacts with the L2 prefetcher and then to understand how explicit prefetching on the L1 also interacts with the L2 prefetcher.

The second objective of the research is related to the interaction of prefetching and virtual/physical memory. On real hardware, prefetching is stopped by page frontiers. The interaction between TLB prefetching (and on which level) and cache prefetching must be analyzed.

The prefetcher is not the only actor in the hierarchy that must be carefully controlled. Significant benefit can also be achieved through careful management of memory access bandwidth, particularly the management of spatial locality on memory accesses, both for reads and writes. The exploitation of this locality is traditionally handled in the memory controller. However, it could be better handled if larger temporal granularity was available. Finally, we also intend to continue to explore the promising avenue of compressed caches. In particular we recently proposed the skewed compressed cache [15]. It offers new possibility for efficient compression schemes.

Ultra wide-issue superscalar.
To effectively leverage memory level parallelism, one requires huge out-of-order execution structures as well as very wide issue superscalar processor. For the two past decades, implementing always wider issue superscalar processor has been challenging. The objective of our research on the execution core is to explore (and revisit) directions to allow the design of a very wide-issue (8-to-16 way) out-of-order execution core while mastering its complexity (silicon area, hardware logic complexity, power/energy consumption).

The first direction that we intend to explore is the use of clustered architecture as in our recent work [8]. Symmetric clustered organization allows to benefit from simpler bypass network, but induce large complexity on the issue queue. One remarkable finding of our study [8] is that, when considering two large clusters (e.g. 8-wide) steering large groups of consecutive instructions (e.g. 64 μops) to the same cluster is quite efficient. This opens opportunities to limit the complexity of the issue queues (monitoring less fewer buses) and register files (reducing number of ports, and number of physical registers) in the clusters, since not all results have to be forwarded to the other cluster.

The second direction that we intend to explore is associated with the approach we developed with Sembrant et al. [16]. It reduces the number of instructions waiting in the instruction queues for the applications benefiting from very large instruction windows. Instructions are dynamically classified as ready (independent from any long latency instruction) or non-ready, and as urgent (part of a dependency chain leading to a long latency instruction) or non-urgent. Non-ready non-urgent instructions can be delayed until the long latency instruction has been executed; this allows to reduce the pressure on the issue queue. This proposition opens the opportunity to consider an asymmetric microarchitecture with a cluster dedicated to the execution of urgent instructions and a second cluster executing the non-urgent instructions. The microarchitecture of this second cluster could be optimized to reduce complexity and power consumption (smaller instruction queue, less aggressive scheduling...)

Speculative execution.
Out-of-order (OoO) execution relies on speculative execution that requires predictions of all sorts: branch, memory dependency, value...

The PACAP members have been major actors of the branch prediction research for the last 20 years; and their proposals have influenced the design of most of the hardware branch predictors in current microprocessors. We will continue to steadily explore new branch predictor designs as for instance [18].

In speculative execution, we have recently revisited value prediction (VP) which was a hot research topic between 1996 and 2002. However it was considered up to recently that value prediction would lead to a huge increase in complexity and power consumption in every stage of the pipeline. Fortunately, we have recently shown that complexity usually introduced by value prediction in the OoO engine can be overcome [12], [11], [19], [9]. First, very high accuracy can be enforced at reasonable cost in coverage and minimal complexity [12]. Thus, both prediction validation and recovery by squashing can be done outside the out-of-order engine, at commit time. Furthermore, we propose a new pipeline organization, EOLE ({Early | Out-of-order | Late} Execution), that leverages VP with validation at commit to execute many instructions outside the OoO core, in-order [11]. With EOLE, the issue-width in OoO core can be reduced without sacrificing performance, thus benefiting the performance of VP without a significant cost in silicon area and/or energy. In the near future, we will explore new avenues related to value prediction. These directions include register equality prediction and compatibility of value prediction with weak memory models in multiprocessors.

3.2.4. Towards heterogeneous single-ISA CPU-GPU architectures

Heterogeneous single-ISA architectures have been proposed in the literature during the 2000's [56] and are now widely used in the industry (ARM big.LITTLE, NVIDIA 4+1...) as a way to improve power-efficiency in mobile processors. These architectures include multiple cores whose respective microarchitectures offer different trade-offs between performance and energy efficiency, or between latency and throughput, while offering the same interface to software. Dynamic task migration policies leverage the heterogeneity of the platform by using the most suitable core for each application, or even each phase of processing. However, these works only tune cores by changing their complexity. Energy-optimized cores are either identical cores implemented in a low-power process technology, or simplified in-order superscalar cores, which are far from state-of-the-art throughput-oriented architectures such as GPUs.

We propose to investigate the convergence of CPU and GPU at both architecture and compilation levels.

Architecture.
The architecture convergence between Single Instruction Multiple Threads (SIMT) GPUs and multicore processors that we have been pursuing [36] opens the way for heterogeneous architectures including latency-optimized superscalar cores and throughput-optimized GPU-style cores, which all share the same instruction set. Using SIMT cores in place of superscalar cores will enable the highest energy efficiency on regular sections of applications. As with existing single-ISA heterogeneous architectures, task migration will not necessitate any software rewrite and will accelerate existing applications.

Compilers for emerging heterogeneous architectures.

Single-ISA CPU+GPU architectures will provide the necessary substrate to enable efficient heterogeneous processing. However, it will also introduce substantial challenges at the software and firmware level. Task placement and migration will require advanced policies that leverage both static information at compile time and dynamic information at run-time. We are tackling the heterogeneous task scheduling problem at the compiler level. As a first step, we are prototyping scheduling algorithms on existing multiple-ISA CPU+GPU architectures like NVIDIA Tegra X1.

3.2.5. Real-time systems

Safety-critical systems (e.g. avionics, medical devices, automotive...) have so far used simple unicore hardware systems as a way to control their predictability, in order to meet timing constraints. Still, many critical embedded systems have increasing demand in computing power, and simple unicore processors are not sufficient anymore. General-purpose multicore processors are not suitable for safety-critical real-time systems, because they include complex micro-architectural elements (cache hierarchies, branch, stride and value predictors) meant to improve average-case performance, and for which worst-case performance is difficult to predict. The prerequisite for calculating tight WCET is a deterministic hardware system that avoids dynamic, time-unpredictable calculations at run-time.

Even for multi and manycore systems designed with time-predictability in mind (Kalray MPPA manycore architecture [0], or the Recore manycore hardware [0]) calculating WCETs is still challenging. The following two challenges will be addressed in the mid-term:

1. definition of methods to estimate WCETs tightly on manycores, that smartly analyzes and/or controls shared resources such as buses, NoCs or caches;
2. methods to improve the programmability of real-time applications through automatic parallelization and optimizations from model-based designs.

3.2.6. Fault Tolerance

Technology trends suggest that, in tomorrow's computing world, failures will become commonplace due to many factors, and the expected probability of failure will increase with scaling. While well-known approaches, such as error correcting codes, exist to recover from failures and provide fault-free chips, the exponential growth of the number of faults will make them unaffordable in the future. Consequently, other approaches such as fine-grained disabling and reconfiguration of hardware elements (e.g. individual functional units or cache blocks) will become economically necessary. We are going to enter a new era: functionally correct chips with variable performance among chips and throughout their lifetime [58].

Transient and permanent faults may be detected by similar techniques, but correcting them generally involves different approaches. We are primarily interested in permanent faults, even though we do not necessarily disregard transient faults (e.g. the TMR approach in the next paragraph addresses both kind of faults).

CPU.
Permanent faults can occur anywhere in the processor. The performance implications of faulty cells vary depending on how the array is used in a processor. Most of micro-architectural work aiming at assessing the performance implications of permanently faulty cells relies on simulations with random fault-maps. These studies are, therefore, limited by the fault-maps they use that may not be representative for the average and distributed performance. They also do not consider aging effect.

Considering the memory hierarchy, we have already studied [5] the impact of permanent faults on the average and worst-case performance based on analytical models. We will extend these models to cover other components and other designs, and to analyze the interaction between faulty components.

For identified critical hardware structures, such as the memory hierarchy, we will propose protection mechanisms by for instance using larger cells, or even by selecting a different array organization to mitigate the impact of faults.

[0] http://www.kalrayinc.com
[0] http://www.recoresystems.com/

Another approach to deal with faults is to introduce redundancy at the code level. We propose to consider static compilation techniques focusing on existing hardware. As an example, we plan to leverage SIMD extensions of current instruction sets to introduce redundancy in scalar code at minimum cost. With these instructions, it will be possible to protect the execution from both soft errors by using TMR (triple modular redundancy) with voters in the code itself, and permanent faults without the need of extra hardware support to deconfigure faulty functional units.

Reconfigurable Computing.
In collaboration with the CAIRN project-team, we propose to construct Coarse Grain Reconfigurable Architectures (CGRA) from a sea of basic arithmetic and memory elements organized into clusters and connected through a hierarchical interconnection network. These clusters of basic arithmetic operators (e.g. 8-bit arithmetic and logic units) would be able to be seamlessly configured to various accuracy and data types to adapt the consumed energy to application requirements taking advantage of approximate computations. We propose to add new kinds of error detection (and sometimes correction) directly at the operator level by taking advantage of the massive redundancy of the array. As an example, errors can be tracked and detected in a complex sequence of double floating-point operations by using a reduced-precision version of the same processing.

Such reconfigurable blocks will be driven by compilation techniques, in charge of computing checkpoints, detecting faults, and replaying computations when needed.

Dynamic compilation techniques will help better exploit faulty hardware, by allocating data and computations on correct resources. In case of permanent faults, we will provide a mechanism to reconfigure the hardware, for example by reducing the issue width of VLIW processors implemented in CGRA. Dynamic code generation (JIT compiler) will re-generate code for the new configuration, guaranteeing portability and optimal exploitation of the hardware.

3.2.7. Power efficiency

PACAP will address power-efficiency at several levels. First, we will design static and split compilation techniques to contribute to the race for Exascale computing (the general goal is to reach 10^{18} FLOP/s at less than 20 MW). Second, we will focus on high-performance low-power embedded compute nodes. will research new static and dynamic compilation techniques that fully exploit emerging memory and NoC technologies. Finally, in collaboration with the CAIRN project-team, we will investigate the synergy of reconfigurable computing and dynamic code generation.

Green and heterogeneous high-performance computing.
Concerning HPC systems, our approach consists in mapping, runtime managing and autotuning applications for green and heterogeneous High-Performance Computing systems up to the Exascale level. One key innovation of the proposed approach consists of introducing a separation of concerns (where self-adaptivity and energy efficient strategies are specified aside to application functionalities) promoted by the definition of a Domain Specific Language (DSL) inspired by aspect-oriented programming concepts for heterogeneous systems. The new DSL will be introduced for expressing adaptivity/energy/performance strategies and to enforce at runtime application autotuning and resource and power management. The goal is to support the parallelism, scalability and adaptability of a dynamic workload by exploiting the full system capabilities (including energy management) for emerging large-scale and extreme-scale systems, while reducing the Total Cost of Ownership (TCO) for companies and public organizations.

High-performance low-power embedded compute nodes.
We will address the design of next generation energy-efficient high-performance embedded compute nodes. It focuses at the same time on software, architecture and emerging memory and communication technologies in order to synergistically exploit their corresponding features. The approach of the project is organized around three complementary topics: 1) compilation techniques; 2) multicore architectures; 3) emerging memory and communication technologies. PACAP will focus on the compilation aspects, taking as input the software-visible characteristics of the proposed emerging technology, and making the best possible use of the new features (non-volatility, density, endurance, low-power).

Hardware Accelerated JIT Compilation.
Reconfigurable hardware offers the opportunity to limit power consumption by dynamically adjusting the number of available resources to the requirements of the running software. In particular, VLIW processors can adjust the number of available issue lanes. Unfortunately, changing the processor width often requires recompiling the application, and VLIW processors are highly dependent of the quality of the compilation, mainly because of the instruction scheduling phase performed by the compiler. Another challenge lies in the high constraints of the embedded system: the energy and execution time overhead due to the JIT compilation must be carefully kept under control.

We started exploring ways to reduce the cost of JIT compilation targeting VLIW-based heterogeneous many-core systems. Our approach lies on a hardware/software JIT compiler framework. While basic optimizations and JIT management are performed in software, the compilation back-end is implemented by means of specialized hardware. This back-end involves both instruction scheduling and register allocation, which are known to be the most time-consuming stages of such a compiler.

3.2.8. Security

Security is a mandatory concern of any modern computing system. Various threat models have led to a multitude of protection solutions. ALF already has contributions, thanks to the HAVEGE [62] random number generator, and code obfuscating techniques (the obfuscating just-in-time compiler [55], or thread-based control flow mangling [60]).

We plan to partner with security experts who can provide intuition, know-how and expertise, in particular in defining threat models, and assessing the quality of the solutions. Our background in compilation and architecture will help design more efficient and less expensive protection mechanisms.

We already have ongoing research directions related to security. We also plan to partner with the Inria/CentraleSupelec CIDRE project-team to design a tainting technique based on a just-in-time compiler.

Compiler-based data protection.
We will specify and design error correction codes suitable for an efficient protection of sensitive information in the context of Internet of Things (IoT) and connected objects. We will partner with experts in security and codes to prototype a platform that demonstrates resilient software. PACAP's expertise will be key to select and tune the protection mechanisms developed within the project, and to propose safe, yet cost-effective solutions from an implementation point of view.

JIT-based tainting.
Dynamic information flow control (DIFC, also known as *tainting*) is used used to detect intrusions and to identify vulnerabilities. It consists in attaching metadata (called *taints* or *labels*) to information containers, and to propagate the taints when particular operations are applied to the containers: reads, writes, etc. The goal is then to guarantee that confidential information is never used to generate data sent to an untrusted container; conversely, data produced by untrusted entities cannot be used to update sensitive data.

The containers can be of various granularities: fine-grain approaches can deal with single variables, coarser-grain approaches consider a file as a whole. The CIDRE project-team has developed several DIFC monitors. kBlare is coarse-grain monitor in the Linux kernel. JBlare is a fine-grain monitor for Java applications. Fine-grain monitors provide a better precision at the cost of a significant overhead in execution time.

We propose to combine the expertise of CIDRE in DIFC with our expertise in JIT compilation to design hybrid approaches. An initial static analysis of the program prior to installation or execution will feed information to a dynamic analyzer that propagates taints during just-in-time compilation.

4. Application Domains

4.1. Any computer usage

The PACAP team is working on the fundamental technologies for computer science: processor architecture, performance-oriented compilation and guaranteed response time for real-time. The research results may have impacts on any application domain that requires high performance execution (telecommunication, multimedia, biology, health, engineering, environment...), but also on many embedded applications that exhibit other constraints such as power consumption, code size and guaranteed response time. Our research activity implies the development of software prototypes.

5. Highlights of the Year

5.1. Highlights of the Year

André Seznec was elevated as an ACM Fellow in December 2016 with the citation: "For contributions to branch prediction and cache memory design".

André Seznec won the three tracks of the 5th Championship on Branch Prediction.

5.1.1. Awards

Sajith Kalathingal, Sylvain Collange, Bharath Swamy and André Seznec received the Best Paper award of the SBAC-PAD 2016 conference.

Damien Hardy, Isabelle Puaut, Yiannakis Sazeides won the best paper award of the Embedded Systems Software track at DATE 2016: Probabilistic WCET estimation in presence of hardware for mitigating the impact of permanent faults. Design, Automation and Test in Europe. Dresden, Germany, March 2016.

Aswinkumar Sridharan and André Seznec won the best paper award for "Discrete Cache Insertion Policies for Shared Last Level Cache Management on Large Multicores" at the 30th IEEE International Parallel & Distributed Processing Symposium, May 2016, Chicago.

For his PhD thesis [10] "Increasing the Performance of Superscalar Processors through Value Prediction", Arthur Perais received:

- Prix de thèse Fondation Rennes 1, 1er Prix de l'école doctorale MATISSE;
- Prix de thèse Gilles Kahn, accessit.

BEST PAPERS AWARDS :

[46] **5th JILP Workshop on Computer Architecture Competitions (JWAC-5): Championship Branch Prediction (CBP-5)**. A. SEZNEC.

[45] **5th JILP Workshop on Computer Architecture Competitions (JWAC-5): Championship Branch Prediction (CBP-5)**. A. SEZNEC.

[36] **International Symposium on Computer Architecture and High-Performance Computing (SBAC-PAD)**. S. KALATHINGAL, S. COLLANGE, B. NARASIMHA SWAMY, A. SEZNEC.

[35] **Design, Automation and Test in Europe**. D. HARDY, I. PUAUT, Y. SAZEIDES.

[48] **30th IEEE International Parallel & Distributed Processing Symposium**. A. SRIDHARAN, A. SEZNEC.

6. New Software and Platforms

6.1. ATMI

KEYWORDS: Analytic model - Chip design - Temperature

SCIENTIFIC DESCRIPTION

Research on temperature-aware computer architecture requires a chip temperature model. General purpose models based on classical numerical methods like finite differences or finite elements are not appropriate for such research, because they are generally too slow for modeling the time-varying thermal behavior of a processing chip.

We have developed an ad hoc temperature model, ATMI (Analytical model of Temperature in MIcroprocessors), for studying thermal behaviors over a time scale ranging from microseconds to several minutes. ATMI is based on an explicit solution to the heat equation and on the principle of superposition. ATMI can model any power density map that can be described as a superposition of rectangle sources, which is appropriate for modeling the microarchitectural units of a microprocessor.

FUNCTIONAL DESCRIPTION

ATMI is a library for modelling steady-state and time-varying temperature in microprocessors. ATMI uses a simplified representation of microprocessor packaging.

- Participant: Pierre Michaud
- Contact: Pierre Michaud
- URL: https://team.inria.fr/pacap/software/atmi/

6.2. Heptane

KEYWORDS: Static analysis - Real time - Performance - WCET - IPET - Worst Case Execution Time

SCIENTIFIC DESCRIPTION

WCET estimation

Status: Registered with APP (Agence de Protection des Programmes). Available under GNU General Public License v3, with number IDDN.FR.001.510039.000.S.P.2003.000.10600.

The aim of Heptane is to produce upper bounds of the execution times of applications. It is targeted at applications with hard real-time requirements (automotive, railway, aerospace domains). Heptane computes WCETs using static analysis at the binary code level. It includes static analyses of microarchitectural elements such as caches and cache hierarchies.

For more information, please contact Damien Hardy or Isabelle Puaut.

FUNCTIONAL DESCRIPTION

In a hard real-time system, it is essential to comply with timing constraints, and Worst Case Execution Time (WCET) in particular. Timing analysis is performed at two levels: analysis of the WCET for each task in isolation taking account of the hardware architecture, and schedulability analysis of all the tasks in the system. Heptane is a static WCET analyser designed to address the first issue.

- Participants: Isabelle Puaut, Damien Hardy, Loïc Besnard
- Partner: Université de Rennes 1
- Contact: Isabelle Puaut
- URL: https://team.inria.fr/pacap/software/heptane/

6.3. Tiptop

KEYWORDS: HPC - Performance - CPU - Cache - Cycles - Instructions - Branch predictor

SCIENTIFIC DESCRIPTION

Tiptop is written in C. It can take advantage of libncurses when available for pseudo-graphic display.

Performance, hardware counters, analysis tool.

Status: Registered with APP (Agence de Protection des Programmes). Available under GNU General Public License v2, with number IDDN.FR.001.450006.000.S.P.2011.000.10800. Current version is 2.3, released July 2015.

Tiptop has been integrated in major Linux distributions, such as Fedora, Debian, Ubuntu.

Tiptop is a new simple and flexible user-level tool that collects hardware counter data on Linux platforms (version 2.6.31+). The goal is to make the collection of performance and bottleneck data as simple as possible, including simple installation and usage. In particular, we stress the following points.

Installation is only a matter of compiling the source code. No patching of the Linux kernel is needed, and no special-purpose module needs to be loaded.

No privilege is required, any user can run tiptop

FUNCTIONAL DESCRIPTION

Today's microprocessors have become extremely complex. To better understand the multitude of internal events, manufacturers have integrated many monitoring counters. Tiptop can be used to collect and display the values from these performance counters very easily. Tiptop may be of interest to anyone who wants to optimise the performance of their HPC applications.

- Participant: Erven Rohou
- Contact: Erven Rohou
- URL: http://tiptop.gforge.inria.fr

6.4. ATC

Address Trace Compression

KEYWORDS: Compressing - Decompressing - Address traces

FUNCTIONAL DESCRIPTION

ATC is a utility and a C library for compressing/decompressing address traces. It implements a new lossless transformation, Bytesort, that exploits spatial locality in address traces. ATC leverages existing general-purpose compressors such as gzip and bzip2. ATC also provides a lossy compression mode that yields higher compression ratios while preserving certain important characteristics of the original trace.

- Participant: Pierre Michaud
- Contact: Pierre Michaud
- URL: https://team.inria.fr/pacap/software/atc/

6.5. Barra

Modelisation of a GPU architecture

KEYWORDS: Simulator - GPU - Computer architecture

SCIENTIFIC DESCRIPTION

Research on throughput-oriented architectures demands accurate and representative models of GPU architectures in order to be able to evaluate new architectural ideas, explore design spaces and characterize applications. The Barra project is a simulator of the NVIDIA Tesla GPU architecture.

Barra builds upon knowledge acquired through micro-benchmarking, in order to provide a baseline model representative of industry practice. The simulator provides detailed statistics to identify optimization opportunities and is fully customizable to experiment ideas of architectural modifications. Barra incorporates both a functional model and a cycle-level performance model.

FUNCTIONAL DESCRIPTION

Barra simulates CUDA programs at the assembly language level (Tesla ISA). Its ultimate goal is to provide a 100 % bit-accurate simulation, offering bug-for-bug compatibility with NVIDIA G80-based GPUs. It works directly with CUDA executables, neither source modification nor recompilation is required.

Barra is primarily intended as a tool for research in computer architecture, although it can also be used to debug, profile and optimize CUDA programs at the lowest level.

- Participants: Sylvain Collange, David Defour, Alexandre Kouyoumdjian and Fabrice Mouhartem
- Contact: Sylvain Collange
- URL: http://barra.gforge.inria.fr/

6.6. If-memo

KEYWORD: Performance, function memoization, dynamic optimization
Status: Ongoing development, early prototype. Registered with APP (Agence de Protection des Programmes) under number IDDN.FR.001.250013.000.S.P.2015.000.10800.

SCIENTIFIC DESCRIPTION

Memoization is the technique of saving result of executions so that future executions can be omitted when the inputs repeat. Memoization has been proposed in previous literature at the instruction level, basic block level and function level using hardware as well as pure software level approaches including changes to programming language.

We proposed software memoization of pure functions for procedural languages. We rely on the operating system loader, taking advantage of the LD_PRELOAD feature of UNIX systems. By setting this variable to the path of a shared library, we instruct the loader to first look to missing symbols in that library. Our library redefines the functions we wish to intercept. The interception code is very straightforward: it receives the same parameter as the target function and checks in a table (a software cache) if this value is readily available. In the favorable case, the result value is immediately returned. Otherwise, we invoke the original function, and store the result in the cache before returning it.

Our technique does not require 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. We validated If-memo with x86-64 platform using both GCC and icc compiler tool-chains, and ARM cortex-A9 platform using GCC.

- Participants: Erven Rohou and Arjun Suresh
- Contact: Erven Rohou

6.7. Padrone

KEYWORDS: Legacy code - Optimization - Performance analysis - Dynamic Optimization
Status: Registered with APP (Agence de Protection des Programmes) under number IDDN.FR.001.250013.000.S.P.2015.000.10800.

FUNCTIONAL DESCRIPTION

Padrone is new platform for dynamic binary analysis and optimization. It provides an API to help clients design and develop analysis and optimization tools for binary executables. Padrone attaches to running applications, only needing the executable binary in memory. No source code or debug information is needed. No application restart is needed either. This is especially interesting for legacy or commercial applications, but also in the context of cloud deployment, where actual hardware is unknown, and other applications competing for hardware resources can vary. The profiling overhead is minimum.

- Participants: Erven Rohou and Emmanuel Riou
- Contact: Erven Rohou
- https://team.inria.fr/pacap/software/Padrone/

6.8. STiMuL

Steady temperature in Multi-Layers components

FUNCTIONAL DESCRIPTION

STiMuL is a C library for modeling steady-state heat conduction in microprocessors. It can be used to obtain temperature from power density or power density from temperature. It can also be used to model stacked dies. STiMuL does not model time-varying temperature. For time-varying temperature, other models must be used, such as ATMI.

- Participant: Pierre Michaud
- Contact: Pierre Michaud
- URL: https://team.inria.fr/pacap/software/stimul/

6.9. TPCalc

Throughput calculator
KEYWORDS: Architecture - Performance analysis
FUNCTIONAL DESCRIPTION

TPCalc is a throughput calculator for microarchitecture studies concerned with multi-program workloads consisting of sequential programs. Because microarchitecture simulators are slow, it is difficult to simulate throughput experiments where a multicore executes many jobs that enter and leave the system. The usual practice of measuring instantaneous throughput on independent coschedules chosen more or less randomly is not a rigorous practice because it assumes that all the coschedules are equally important, which is not always true. TPCalc can compute the average throughput of a throughput experiment without actually doing the throughput experiment. The user first defines the workload heterogeneity (number of different job types), the multicore configuration (number of cores and symmetries). TPCalc provides a list of base coschedules. The user then simulates these coschedules, using some benchmarks of his choice, and feeds back to TPCalc the measured execution rates (e.g., instructions per cycle or instructions per second).TPCalc eventually outputs the average throughput.

- Participant: Pierre Michaud
- Partner: Ghent University
- Contact: Pierre Michaud
- URL: https://team.inria.fr/pacap/software/tpcalc/

6.10. Parasuite

Participants: Sylvain Collange, Imane Lasri, Erven Rohou, André Seznec.

Parasuite: parallel benchmarks for multi-core CPUs, clusters and accelerators

Despite the ubiquity of parallel architectures in all computing segments, the research community often lacks benchmarks representative of parallel applications. The Inria Parallel Benchmark Suite (Parasuite) seeks to address this need by providing a set of representative parallel benchmarks for the architecture, compiler and system research communities. Parasuite targets the main contemporary parallel programming technologies: shared-memory multi-thread parallelism for multi-core, message-passing parallelism for clusters and fine-grained data-level parallelism for GPU architectures and SIMD extensions.

All benchmarks come with input datasets of various sizes, to accommodate use cases ranging from microarchitecture simulation to large-scale performance evaluation. Correctness checks on the computed results enable automated regression testing. In order to support computer arithmetic optimization and approximate computing research scenarios, the correctness checks favor accuracy metrics evaluating domain-specific relevance rather than bit-exact comparisons against an arbitrary reference output.

Visit: http://parasuite.inria.fr/

6.11. Simty

Participant: Sylvain Collange.

Simty: A Synthesizable General-Purpose SIMT Processor.

Simty is a massively multi-threaded processor core that dynamically assembles SIMD instructions from scalar multi-thread code. It runs the RISC-V (RV32-I) instruction set. Unlike existing SIMD or SIMT processors like GPUs, Simty takes binaries compiled for general-purpose processors without any instruction set extension or compiler changes. Simty is described in synthesizable VHDL.

Visit: http://team.inria.fr/pacap/simty

7. New Results

7.1. Compiler, vectorization, interpretation

Participants: Erven Rohou, Emmanuel Riou, Arjun Suresh, André Seznec, Nabil Hallou, Sylvain Collange, Rabab Bouziane, Arif Ali Ana-Pparakkal, Stefano Cherubin.

7.1.1. Improving sequential performance through memoization

Participants: Erven Rohou, Emmanuel Riou, 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 propose [20] 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.

Initial work has been published in ACM TACO 2015 [20] and accepted for presentation at the International Conference HiPEAC 2016 in Prague.

Further developments have been accepted for publication at the Compiler Construction Conference 2017 [49].

This research is described in the PhD thesis of Arjun Suresh [24].

7.1.2. Optimization in the Presence of NVRAM

Participants: Erven Rohou, Rabab Bouziane.

Energy-efficiency is one of the most challenging design issues in both embedded and high-performance computing domains. The aim is to reduce as much as possible the energy consumption of considered systems while providing them with the best computing performance. Finding an adequate solution to this problem certainly requires a cross-disciplinary approach capable of addressing the energy/performance trade-off at different system design levels.

We proposed [42] an empirical impact analysis of the integration of Spin Transfer Torque Magnetic Random Access Memory (STT-MRAM) technologies in multicore architectures when applying some existing compiler optimizations. For that purpose, we use three well-established architecture and NVM evaluation tools: NVSim, gem5 and McPAT. Our results show that the integration of STT-MRAM at cache memory levels enables a significant reduction of the energy consumption (up to 24.2 % and 31 % on the considered multicore and monocore platforms respectively) while preserving the performance improvement provided by typical code optimizations. We also identified how the choice of the clock frequency impacts the relative efficiency of the considered memory technologies.

This research is done in collaboration with Abdoulaye Gamatié at LIRMM (Montpellier) within the context the the ANR project CONTINUUM.

7.1.3. Hardware/Software JIT Compiler

Participant: Erven Rohou.

Dynamic Binary Translation (DBT) is often used in hardware/software co-design to take advantage of an architecture model while using binaries from another one. The co-development of the DBT engine and of the execution architecture leads to architecture with special support to these mechanisms. We proposed a hardware accelerated dynamic binary translation where the first steps of the DBT process are fully accelerated in hardware. Results shows that using our hardware accelerators leads to a speed-up of $8\times$ and a cost in energy $18\times$ lower, compared with an equivalent software approach.

An initial version of this work has been presented at Compas'16 [51]. The latest results have been accepted for publication at DATE 2017 [44].

This research is done in collaboration with Steven Derrien and Simon Rokicki from the CAIRN team.

7.1.4. Dynamic Parallelization of Binary Programs

Participants: Erven Rohou, Emmanuel Riou, Nabil Hallou.

We address runtime automatic parallelization of binary executables, assuming no previous knowledge on the executable code. The Padrone platform is used to identify candidate functions and loops. Then we disassemble the loops and convert them to the intermediate representation of the LLVM compiler. This allows us to leverage the power of the polyhedral model for auto-parallelizing loops. Once optimized, new native code is generated just-in-time in the address space of the target process.

Our approach enables user transparent auto-parallelization of legacy and/or commercial applications with auto-parallelization.

This work has been accepted for publication in the Springer journal IJPP: "Runtime Vectorization Transformations of Binary Code".

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

7.1.5. Dynamic Function Specialization

Participants: Erven Rohou, Arif Ali Ana-Pparakkal.

Compilers can do better optimization with the knowledge of run-time behaviour of the program. *Function Specialization* is an optimization technique in which different versions of a function are created according to the value of its arguments. It can be difficult to predict the exact value/behaviour of arguments during static compilation and so it is difficult for a static compiler to do efficient function specialization. In our *dynamic function specialization* technique, we capture the actual value of arguments during execution of the program and, when profitable, create specialized versions and include them at runtime.

This research is done within the context of the Nano 2017 PSAIC collaborative project.

7.1.6. Application Autotuning for Performance and Energy

Participants: Erven Rohou, Stefano Cherubin, Imane Lasri.

Due to the increasing complexity of both applications behaviors and underlying hardware, achieving reasonable (not to mention best) performance can hardly be done at compile time. Autotuning is an approach where a runtime manager is able to adapt the software to the runtime conditions. We have developed a framework and shown through a domain specific application initial exploration scenarios [32], [47].

We started characterizing applications – in particular the Parasuite benchmarks – and we will rely on split-compilation [2] embed hints and heuristics inside a binary program for dynamic adaptation and optimization.

This research is done within the context of the H2020 FET HPC collaborative project ANTAREX.

7.1.7. Customized Precision Computing

Participants: Erven Rohou, Stefano Cherubin, Imane Lasri.

Customized precision originates from the fact that many applications can tolerate some loss of quality during computation, as in the case of media processing (audio, video and image), data mining, machine learning, etc. Error-tolerating applications are increasingly common in the emerging field of real-time HPC. Thus, recent works have investigated this line of research in the HPC domain as a way to provide a breakthrough in power and performance for the Exascale era.

We aim at leveraging existing, HPC-oriented hardware architectures, while including in the precision tuning an adaptive selection of floating and fixed-point arithmetic. It is part of a wider effort to provide the programmers with an easy way to manage extra-functional properties of programs, including precision, power, and performance.

We explore tradeoffs between precision and time-to-solution, as well as precision and energy-to-solution.

This is done within the context of the ANTAREX project in collaboration with Stefano Cherubin, Cristina Silvano and Giovanni Agosta from Politecnico di Milano, and Olivier Sentieys from the CAIRN team.

7.1.8. SPMD Function Call Re-Vectorization

Participant: Sylvain Collange.

SPMD programming languages for SIMD hardware such as C for CUDA, OpenCL or ISPC have contributed to increase the programmability of SIMD accelerators and graphics processing units. However, SPMD languages still lack the flexibility offered by low-level SIMD programming on explicit vectors. To close this expressiveness gap while preserving the SPMD abstraction, we introduce the notion of Function Call Re-Vectorization (CREV) [38]. CREV allows changing the dimension of vectorization during the execution of an SPMD kernel, and exposes it as a nested parallel kernel call. CREV affords a programmability close to dynamic parallelism, a feature that allows the invocation of kernels from inside kernels, but at much lower cost. In this paper, we present a formal semantics of CREV, and an implementation of it on the ISPC compiler. To validate our idea, we have used CREV to implement some classic algorithms, including string matching, depth first search and Bellman-Ford, with minimum effort. These algorithms, once compiled by ISPC to Intel-based vector instructions, are as fast as state-of-the-art implementations, yet much simpler. As an example, our straightforward implementation of string matching beats the Knuth-Morris-Pratt algorithm by 12 %.

This work was done during the internship of Rubens Emilio in Rennes in collaboration with Sylvain Collange and Fernando Pereira (UFMG) as part of the Inria PROSPIEL Associate Team.

7.1.9. SPMD Function Call Fusion

Participant: Sylvain Collange.

The increasing popularity of Graphics Processing Units (GPUs) has brought renewed attention to old problems related to the Single Instruction, Multiple Data execution model. One of these problems is the reconvergence of divergent threads. A divergence happens at a conditional branch when different threads disagree on the path to follow upon reaching this split point. Divergences may impose a heavy burden on the performance of parallel programs.

We have proposed a compiler-level optimization to mitigate the performance loss due to branch divergence on GPUs [21]. This optimization consists in merging function call sites located at different paths that sprout from the same branch. We show that our optimization adds negligible overhead on the compiler. When not applicable, it does not slow down programs and it accelerates substantially those in which it is applicable. As an example, we have been able to speed up the well known SPLASH Fast Fourier Transform benchmark by 11 %.

This work is done in collaboration with Douglas do Couto Teixeira and Fernando Pereira from UFMG as part of the Inria PROSPIEL Associate Team.

7.1.10. SIMD programming in SPMD: application to multi-precision computations

Participant: Sylvain Collange.

GPUs are an important hardware development platform for problems where massive parallel computations are needed. Many of these problems require a higher precision than the standard double floating-point (FP) available. One common way of extending the precision is the multiple-component approach, in which real numbers are represented as the unevaluated sum of several standard machine precision FP numbers. This representation is called a FP expansion and it offers the simplicity of using directly available and highly optimized FP operations. We propose new data-parallel algorithms for adding and multiplying FP expansions specially designed for extended precision computations on GPUs [34]. These are generalized algorithms that can manipulate FP expansions of different sizes (from double-double up to a few tens of doubles) and ensure a certain worst case error bound on the results.

This work is done in collaboration with Mioara Joldes (CNRS/LAAS), Jean-Michel Muller (CNRS/LIP) and Valentina Popescu (ENS Lyon/LIP).

7.2. Processor Architecture

Participants: Pierre Michaud, Sylvain Collange, Erven Rohou, André Seznec, Arthur Perais, Sajith Kalathingal, Andrea Mondelli, Aswinkumar Sridharan, Biswabandan Panda, Fernando Endo, Kleovoulos Kalaitzidis.

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

7.2.1. Microarchitecture

7.2.1.1. Branch prediction

Participant: André Seznec.

IMLI-based predictors

The wormhole (WH) branch predictor was recently introduced to exploit branch outcome correlation in multidimensional loops. For some branches encapsulated in a multidimensional loop, their outcomes are correlated with those of the same branch in neighbor iterations, but in the previous outer loop iteration. In [18], we introduced practical predictor components to exploit this branch outcome correlation in multidimensional loops: the IMLI-based predictor components. The iteration index of the inner most loop in an application can be efficiently monitored at instruction fetch time using the Inner Most Loop Iteration (IMLI) counter. The outcomes of some branches are strongly correlated with the value of this IMLI counter. Our experiments show that augmenting a state-of-the-art global history predictor such as TAGE-SC-L [45] with IMLI-based components outperforms previous state-of-the-art academic predictors leveraging local and global history at much lower hardware complexity (i.e., smaller storage budget , smaller number of tables and simpler management of speculative states).

This study was accepted in the special issue Top Picks of the best papers in 2015 computer architecture conferences in IEEE Micro [30].

This research was done in collaboration with Joshua San Miguel and Jorge Albericio from University of Toronto

Championship Branch Prediction

The 5th Championship Branch Prediction was organized in Seoul in June 2016. The predictors submitted by the PACAP-team, respectively TAGE-SC-L and MTAGE-SC, for limited storage budgets and infinite storage budgets won the three tracks of the competition [46], [45]. These predictors are derived from our reference work [17].

7.2.1.2. 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. From 2013 to 2016, we have progressively revived the interest in value prediction. At a first step, we showed that all predictors are amenable to very high accuracy at the cost of some loss on prediction coverage [12]. Furthermore, we proposed EOLE [13]. EOLE leverages Value Prediction to *Early Execute* simple instructions whose operands are ready in parallel with Rename and to *Late Execute* to simple predicted instructions just before Commit. EOLE allows to reduce the out-of-order issue-width by 33% without impeding performance.

An extension of the initial EOLE paper [13] was published in ACM TOCS [27].

7.2.1.3. Physical register sharing

Participants: Arthur Perais, André Seznec.

Sharing a physical register between several instructions is needed to implement several microarchitectural optimizations. However, register sharing requires modifications to the register reclaiming process: Committing a single instruction does not guarantee that the physical register allocated to the previous mapping of its architectural destination register is free-able anymore. Consequently, a form of register reference counting must be implemented. While such mechanisms (e.g., dependency matrix, per register counters) have been described in the literature, we argue that they either require too much storage, or that they lengthen branch misprediction recovery by requiring sequential rollback. As an alternative, we present the Inflight Shared Register Buffer (ISRB), a new structure for register reference counting [41]. The ISRB has low storage overhead and lends itself to checkpoint-based recovery schemes, therefore allowing fast recovery on pipeline flushes. We illustrate our scheme with Move Elimination (short-circuiting moves) and an implementation of Speculative Memory Bypassing (short-circuiting store-load pairs) that makes use of a TAGE-like predictor to identify memory dependencies. We show that the whole potential of these two mechanisms can be achieved with a small register tracking structure.

7.2.1.4. Register Sharing for Equality Prediction

Participants: Arthur Perais, Fernando Endo, André Seznec.

Recently, Value Prediction (VP) has been gaining renewed traction in the research community. VP speculates on the result of instructions to increase Instruction Level Parallelism (ILP). In most embodiments, VP requires large tables to track predictions for many static instructions. However, in many cases, it is possible to detect that the result of an instruction is produced by an older inflight instruction, but not to predict the result itself. Consequently it is possible to rely on predicting register equality and handle speculation through the renamer. To do so, we propose to use Distance Prediction [40], a technique that was previously used to perform Speculative Memory Bypassing (short-circuiting def-store-load-use chains). Distance Prediction attempts to determine how many instructions separate the instruction of interest and the most recent older instruction that produced the same result. With this information, the physical register identifier of the older instruction can be retrieved from the ROB and provided to the renamer. The implementation of Distance Prediction necessitates a hardware mechanism to handle the sharing of physical registers as the ISRB [41].

7.2.1.5. Storage-Free Memory Dependency Prediction

Participants: Arthur Perais, André Seznec.

Memory Dependency Prediction (MDP) is paramount to good out-of-order performance, but decidedly not trivial as a all instances of a given static load may not necessarily depend on all instances of a given static store. As a result, for a given load, MDP should predict the exact store instruction the load depends on, and not only whether it depends on an inflight store or not, i.e., ideally, prediction should not be binary. However, we first argue that given the high degree of sophistication of modern branch predictors, the fact that a given dynamic load depends on an inflight store can be captured using the binary prediction capabilities of the branch predictor, providing coarse MDP at zero storage overhead. Second, by leveraging hysteresis counters, we show that the precise producer store can in fact be identified. This embodiment of MDP yields performance levels that are on par with state-of-the-art, and requires less than 70 additional bits of storage over a baseline without MDP at all [28].

7.2.1.6. Compressed Caches

Participants: André Seznec, Biswabandan Panda.

The YACC compressed cache

Cache memories play a critical role in bridging the latency, bandwidth, and energy gaps between cores and off-chip memory. However, caches frequently consume a significant fraction of a multicore chip's area, and thus account for a significant fraction of its cost. Compression has the potential to improve the effective capacity of a cache, providing the performance and energy benefits of a larger cache while using less area. The design of a compressed cache must address two important issues: i) a low-latency, low-overhead compression algorithm that can represent a fixed-size cache block using fewer bits and ii) a cache organization that can efficiently store the resulting variable-size compressed blocks. This paper focuses on the latter issue. We propose YACC (Yet Another Compressed Cache), a new compressed cache design that targets improving effective cache capacity with a simple design [29]. YACC uses super-blocks to reduce tag overheads, while packing variable-size compressed blocks to reduce internal fragmentation. YACC achieves the benefits of two state-of-the art compressed caches, Decoupled Compressed Cache (DCC) [61] and Skewed Compressed Cache (SCC) [15], with a more practical and simpler design. YACC's cache layout is similar to conventional caches, with a largely unmodified tag array and unmodified data array.

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

The DISH compression scheme

The effectiveness of a compressed cache depends on three features: i) the compression scheme, ii) the compaction scheme, and iii) the cache layout of the compressed cache. Both SCC [15] and YACC [29] use compression techniques to compress individual cache blocks, and then a compaction technique to compact multiple contiguous compressed blocks into a single data entry. The primary attribute used by these techniques for compaction is the compression factor of the cache blocks, and in this process, they waste cache space. We propose dictionary sharing (DISH), a dictionary based cache compression scheme that reduces this wastage [39]. DISH compresses a cache block by keeping in mind that the block is a potential candidate for the compaction process. DISH encodes a cache block with a dictionary that stores the distinct 4-byte chunks of a cache block and the dictionary is shared among multiple neighboring cache blocks. The simple encoding scheme of DISH also provides a single cycle decompression latency and it does not change the cache layout of compressed caches. Compressed cache layouts that use DISH outperforms the compression schemes, such as BDI and CPACK+Z, in terms of compression ratio, system performance , and energy efficiency.

7.2.1.7. 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.

In [8], 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 study published in ACM TACO in 2015 [8] and was presented at the HIPEAC 2016 conference.

7.2.1.8. Hardware data prefetching

Participant: Pierre Michaud.

Hardware prefetching is an important feature of modern high-performance processors. When an application's working set is too large to fit in on-chip caches, disabling hardware prefetchers may result in severe performance reduction. We propose a new hardware data prefetcher, the Best-Offset (BO) prefetcher. The BO prefetcher is an offset prefetcher using a new method for selecting the best prefetch offset taking into account prefetch timeliness. The hardware required for implementing the BO prefetcher is very simple. A version of the BO prefetcher won the 2015 Data Prefetching Championship. A comprehensive study of the BO prefetcher was presented at the HPCA 2016 conference [37].

7.2.1.9. Exploiting loops for lower energy consumption

Participants: Andrea Mondelli, Pierre Michaud, André Seznec.

Recent superscalar processors use a loop buffer to decrease the energy consumption in the front-end. The energy savings comes from the branch predictor, instruction cache and instruction decoder being idle when micro-ops are delivered to the back-end from the loop buffer. We explored the possibility to exploit loop behaviors for decreasing energy consumption further, in the back-end, without impacting performance. We proposed two independent optimizations requiring little extra hardware. The first optimization detects and removes from the execution redundant micro-ops producing the same result in every loop iteration. The second optimization focuses on loop loads and detects situations where a loop load needs accessing only the data cache, or only the store queue, not both.

7.2.2. Microarchitecture Performance Modeling

7.2.2.1. Optimal cache replacement

Participant: Pierre Michaud.

A cache replacement policy is an algorithm, implemented in hardware, selecting a block to evict to make room for an incoming block. This research topic has been revitalized recently, as level-2 and level-3 caches were integrated on chip. A cache replacement policy cannot be optimal in general unless it has the knowledge of future references. Unfortunately, practical replacement policies do not have this knowledge. Still, optimal replacement is an important benchmark for understanding replacement policies. Moreover, some new replacement policies proposed recently are directly inspired from algorithms for determining hits and misses under optimal replacement. Hence it is important to improve our understanding of optimal replacement.

The OPT policy, which evicts the block referenced furthest in the future, was proved optimal by Mattson et al. [57]. However, their proof is long and somewhat complicated. In collaboration with some researchers from Inha University, we found a shorter and more intuitive proof of optimality for OPT [6].

An intriguing aspect of optimal replacement, seldom mentioned in the literature, is the fact that Belady's MIN algorithm determines OPT hits and misses without the knowledge of future references [54]. Starting from this fact, we searched and found a new algorithm, different from MIN, for determining OPT hits and misses. This algorithm provides new insights about optimal replacement. We show that traces of OPT stack distances have a distinctive structure. In particular, we prove that OPT miss curves are always convex. We show that, like an LRU cache, an OPT cache cannot experience more misses as the reuse distance of references is decreased. Consequently, accessing data circularly is the worst access pattern for OPT, like it is for LRU. We discovered an equivalence between an OPT cache of associativity N with bypassing allowed and an OPT cache of associativity N+1 with bypassing disabled. A paper deriving these results was accepted in ACM TACO and will be presented at the HiPEAC 2017 conference [25].

7.2.2.2. Adaptive Intelligent Memory Systems

Participants: André Seznec, Aswinkumar Sridharan.

Multi-core processors employ shared Last Level Caches (LLC). This trend will continue in the future with large multi-core processors (16 cores and beyond) as well. At the same time, the associativity of this LLC tends to remain in the order of sixteen. Consequently, with large multicore processors, the number of cores that share the LLC becomes larger than the associativity of the cache itself. LLC management policies have been extensively studied for small scale multi-cores (4 to 8 cores) and associativity degree in the 16 range. However, the impact of LLC management on large multi-cores is essentially unknown, in particular when the associativity degree is smaller than the number of cores.

In [48], we introduce Adaptive Discrete and deprioritized Application PrioriTization (ADAPT), an LLC management policy addressing the large multi-cores where the LLC associativity degree is smaller than the number of cores. ADAPT builds on the use of the Footprint-number metric. Footprint-number is defined as the number of unique accesses (block addresses) that an application generates to a cache set in an interval of time. We propose a monitoring mechanism that dynamically samples cache sets to estimate the Footprint-number of applications and classifies them into discrete (distinct and more than two) priority buckets. The cache replacement policy leverages this classification and assigns priorities to cache lines of applications during cache replacement operations. Footprint-number is computed periodically to account the dynamic changes in applications behavior. We further find that de- prioritizing certain applications during cache replacement is beneficial to the overall performance. We evaluate our proposal on 16, 20 and 24-core multi-programmed workloads and discuss other aspects in detail.

[48] got the best paper award at the IPDPS 2016 conference.

7.2.2.3. Augmenting superscalar architecture for efficient many-thread parallel execution

Participants: Sylvain Collange, André Seznec, Sajith Kalathingal.

Threads of Single-Program Multiple-Data (SPMD) applications often exhibit very similar control flows, i.e. they execute the same instructions on different data. In [36] we propose the Dynamic Inter-Thread Vectorization Architecture (DITVA) to leverage this implicit data-level parallelism in SPMD applications by assembling dynamic vector instructions at runtime. DITVA extends an in-order SMT processor with SIMD units with an inter-thread vectorization execution mode. In this mode, multiple scalar threads running in lockstep share a single instruction stream and their respective instruction instances are aggregated into SIMD instructions. To balance thread-and data-level parallelism, threads are statically grouped into fixed-size independently scheduled warps. DITVA leverages existing SIMD units and maintains binary compatibility with existing CPU architectures. Our evaluation on the SPMD applications from the PARSEC and Rodinia OpenMP benchmarks shows that a 4-warp $\times$ 4-lane 4-issue DITVA architecture with a realistic bank-interleaved cache achieves 1.55$\times$ higher performance than a 4-thread 4-issue SMT architecture with AVX instructions while fetching and issuing 51 % fewer instructions, achieving an overall 24 % energy reduction.

Our paper [36] received the Best Paper Award of the SBAC-PAD conference.

7.2.2.4. *Generalizing the SIMT execution model to general-purpose instruction sets*

Participant: Sylvain Collange.

The *Single Instruction, Multiple Threads* (SIMT) execution model as implemented in NVIDIA Graphics Processing Units (GPUs) associates a multi-thread programming model with an SIMD execution model [59]. It combines the simplicity of scalar code from the programmer's and compiler's perspective with the efficiency of SIMD execution units at the hardware level. However, current SIMT architectures demand specific instruction sets. In particular, they need specific branch instructions to manage thread divergence and convergence. Thus, SIMT GPUs have remained incompatible with traditional general-purpose CPU instruction sets.

We designed Simty, an SIMT processor proof of concept that lifts the instruction set incompatibility between CPUs and GPUs [50]. Simty is a massively multi-threaded processor core that dynamically assembles SIMD instructions from scalar multi-thread code. It runs the RISC-V (RV32-I) instruction set. Unlike existing SIMD or SIMT processors like GPUs, Simty takes binaries compiled for general-purpose processors without any instruction set extension or compiler changes. Simty is described in synthesizable RTL. A FPGA prototype validates its scaling up to 2048 threads per core with 32-wide SIMD units.

7.3. WCET estimation and optimization

Participants: Isabelle Puaut, Damien Hardy, Viet Anh Nguyen, Benjamin Rouxel, Sébastien Martinez, Erven Rohou.

7.3.1. WCET estimation for many core processors

Participants: Viet Anh Nguyen, Damien Hardy, Sébastien Martinez, Isabelle Puaut, Benjamin Rouxel.

7.3.1.1. *Optimization of WCETs by considering the effects of local caches*

The overall goal of this research is to defined WCET estimation methods for parallel applications running on many-core architectures, such as the Kalray MPPA machine.

Some approaches to reach this goal have been proposed, but they assume the mapping of parallel applications on cores already done. Unfortunately, on architectures with caches, task mapping requires a priori known WCETs for tasks, which in turn requires knowing task mapping (i.e., co-located tasks, co-running tasks) to have tight WCET bounds. Therefore, scheduling parallel applications and estimating their WCET introduce a chicken and egg situation.

We address this issue by developing both optimal and heuristic techniques for solving the scheduling problem, whose objective is to minimize the WCET of a parallel application. Our proposed static partitioned non-preemptive mapping strategies address the effect of local caches to tighten the estimated WCET of the parallel application. Experimental results obtained on real and synthetic parallel applications show that co-locating tasks that reuse code and data improves the WCET.

This research is part of the PIA Capacités project.

7.3.1.2. *Accounting for shared resource contentions to minimize WCETs*

Accurate WCET analysis for multi-cores is known to be challenging, because of concurrent accesses to shared resources, such as communication through busses or Networks on Chips (NoC). Since it is impossible in general to guarantee the absence of resource conflicts during execution, current WCET techniques either produce pessimistic WCET estimates or constrain the execution to enforce the absence of conflicts, at the price of a significant hardware under-utilization. In addition, the large majority of existing works consider that the platform workload consists of independent tasks. As parallel programming is the most promising solution to improve performance, we envision that within only a few years from now, real-time workloads will evolve toward parallel programs. The WCET behavior of such programs is challenging to analyze because they consist of *dependent* tasks interacting through complex synchronization/communication mechanisms.

In this work, we propose techniques that account for interferences to access shared ressources, in order to minimize the WCET of parallel applications. An optimal and a heuristic method are proposed to map and schedule tasks on multi-cores. These methods take the structure of applications (synchronizations/communications) into consideration to tightly identify shared resource interferences and consequently tighten WCET estimates.

This work is performed in cooperation with Steven Derrien, Angeliki Kritikakou and Imen Fassi from the CAIRN research group and is part of the ARGO H2020 project.

7.3.2. Cache-Persistence-Aware Response-Time Analysis for Fixed-Priority Preemptive Systems

Participants: Damien Hardy, Isabelle Puaut.

A task can be preempted by several jobs of higher priority tasks during its execution. Assuming the worst-case memory demand for each of these jobs leads to pessimistic worst-case response time (WCRT) estimations. Indeed, there is a big chance that a large portion of the instructions and data associated with the preempting task τ_j are still available in the cache when τ_j releases its next jobs. Accounting for this observation allows the pessimism of WCRT analysis to be significantly reduced, which is not considered by existing work.

The four main contributions of this work are: 1) The concept of persistent cache blocks is introduced in the context of WCRT analysis, which allows re-use of cache blocks to be captured, 2) A cache-persistence-aware WCRT analysis for fixed-priority preemptive systems exploiting the PCBs to reduce the WCRT bound, 3) A multi-set extension of the analysis that further improves the WCRT bound and 4) An evaluation showing that our cache-persistence-aware WCRT analysis results in up to 10 % higher schedulability than state-of-the-art approaches.

This work [43] appeared at ECRTS 2016 and was selected as an outstanding paper in this conference.

This work was performed in cooperation with Syed Aftab Rashid, Geoffrey Nelissen, Benny Akesson and Eduardo Tovar from ISEP (Polytechnic Institute of Porto), Portugal.

7.4. Fault Tolerance

7.4.1. WCET estimation for architectures with faulty caches

Participants: Damien Hardy, Isabelle Puaut.

Fine-grained disabling and reconfiguration of hardware elements (functional units, cache blocks) will become economically necessary to recover from permanent failures, whose rate is expected to increase dramatically in the near future. This fine-grained disabling will lead to degraded performance as compared to a fault-free execution.

Until recently, all static worst-case execution time (WCET) estimation methods were assuming fault-free processors, resulting in unsafe estimates in the presence of faults. The first static WCET estimation technique dealing with the presence of permanent faults in instruction caches was proposed in [4]. This study probabilistically quantified the impact of permanent faults on WCET estimates. It demonstrated that the probabilistic WCET (pWCET) estimates of tasks increase rapidly with the probability of faults as compared to fault-free WCET estimates.

New results show that very simple reliability mechanisms allow mitigating the impact of faulty cache blocks on pWCETs. Two mechanisms, that make part of the cache resilient to faults are analyzed. Experiments show that the gain in pWCET for these two mechanisms are on average 48 % and 40 % as compared to an architecture with no reliability mechanism.

This work [35] appeared at DATE 2016 (best paper award for the embedded systems track).

This is joint work with Yannakis Sazeides from University of Cyprus.

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Contracts with Industry

8.1.1. Nano 2017 PSAIC

Participants: Arif Ali Ana-Pparakkal, Erven Rohou, Emmanuel Riou.

Nano 2017 PSAIC is a collaborative R&D program involving Inria and STMicroelectronics. The PSAIC (Performance and Size Auto-tuning through Iterative Compilation) project concerns the automation of program optimization through the combination of several tools and techniques such as: compiler optimization, profiling, trace analysis, iterative optimization and binary analysis/rewriting. For any given application, the objective is to devise through a fully automated process a compiler profile optimized for performance and code size. For this purpose, we are developing instrumentation techniques that can be focused and specialized to a specific part of the application aimed to be monitored.

The project involves the Inria teams PACAP, AriC, CAMUS and CORSE. PACAP contributes program analyses at the binary level, as well as binary transformations. We will also study the synergy between static (compiler-level) and dynamic (run-time) analyses.

8.2. Bilateral Grants with Industry

8.2.1. Intel research grant INTEL2014-8957

Participants: André Seznec, Biswabandan Panda, Arthur Perais, Fernando Endo.

Intel is supporting the research of the PACAP project-team on "Mixing branch and value prediction to enable high sequential performance".

8.2.2. Intel research grant INTEL2016-11174

Participants: André Seznec, Pierre Michaud, Kleovoulos Kalaitzidis.

Intel is supporting the research of the PACAP project-team on "Design tradeoffs for extreme cores".

9. Partnerships and Cooperations

9.1. National Initiatives

9.1.1. Capacités: Projet "Investissement d'Avenir", 1/11/14 to 31/01/2018

Participants: Damien Hardy, Isabelle Puaut, Viet Anh Nguyen, Sébastien Martinez.

The project objective is to develop a hardware and software platform based on manycore architectures, and to demonstrate the relevance of these manycore architectures (and more specifically the Kalray manycore) for several industrial applications. The Kalray MPPA manycore architecture is currently the only one able to meet the needs of embedded systems simultaneously requiring high performance, lower power consumption, and the ability to meet the requirements of critical systems (low latency I/O, deterministic processing times, and dependability). The project partners are Kalray (lead), Airbus, Open-Wide, Safran Sagem, IS2T, Real Time at Work, Dassault Aviation, Eurocopter, MBDA, ProbaYes, IRIT, Onera, Verimag, Inria, Irisa, Tima and Armines.

9.1.2. Multicore: Inria Project Lab, 2013-2016

Participants: Erven Rohou, Nabil Hallou.

Multicore is an Inria Project Lab (IPL, formerly *Action d'Envergure*) started in 2013. It is entitled "Large scale multicore virtualization for performance scaling and portability". Partner project-teams include: PACAP, ALGORILLE, CAMUS, REGAL, RUNTIME, as well as DALI. This project aims to build collaborative virtualization mechanisms that achieve essential tasks related to parallel execution and data management. We want to unify the analysis and transformation processes of programs and accompanying data into one unique virtual machine.

9.1.3. ANR Continuum 2015–2019

Participants: Erven Rohou, Rabab Bouziane.

The CONTINUUM project aims to address the energy-efficiency challenge in future computing systems by investigating a design continuum for compute nodes, which seamlessly goes from software to technology levels via hardware architecture. Power saving opportunities exist at each of these levels, but the real measurable gains will come from the synergistic focus on all these levels as considered in this project. Then, a cross-disciplinary collaboration is promoted between computer science and microelectronics, to achieve two main breakthroughs: i) combination of state-of-the-art heterogeneous adaptive embedded multicore architectures with emerging communication and memory technologies and, ii) power-aware dynamic compilation techniques that suitably match such a platform.

Continuum started on Oct 1st 2015. Partners are LIRMM and Cortus SAS.

9.1.4. ANR CHIST-ERA SECODE 2016-2018

Participants: Nicolas Kiss, Damien Hardy, Erven Rohou.

In this project, we specify and design error correction codes suitable for an efficient protection of sensitive information in the context of Internet of Things (IoT) and connected objects. Such codes mitigate passive attacks, like memory disclosure, and active attacks, like stack smashing. The innovation of this project is to leverage these codes for protecting against both cyber and physical attacks. The main advantage is a full coverage of attacks of the connected embedded systems, which is considered as a smart connected device and also a physical device. The outcome of the project is first a method to generate and execute cyber-resilient software, and second to protect data and its manipulation from physical threats like side-channel attacks. Theses results are demonstrated by using a smart sensor application with hardened embedded firmware and tamper-proof hardware platform.

Partners are Télécom Paris Tech, Université Paris 8, University of Sabancı(Turkey), and Université Catholique de Louvain (Belgium).

9.1.5. ANR W-SEPT 2012-2016

Participants: Isabelle Puaut, Erven Rohou.

Critical embedded systems are generally composed of repetitive tasks that must meet drastic timing constraints, such as termination deadlines. Providing an upper bound of the worst-case execution time (WCET) of such tasks at design time is thus necessary to prove the correctness of the system. Static WCET estimation methods, although safe, may produce largely over-estimated values. The objective of the project is to produce tighter WCET estimates by discovering and transforming flow information at all levels of the software design process, from high level-design models (e.g. Scade, Simulink) down to binary code. The ANR W-SEPT project partners are Verimag Grenoble, IRIT Toulouse, Inria Rennes. A case study is provided by Continental Toulouse.

9.1.6. PEPS INS2I gDGA

Participant: Sylvain Collange.

This interdisciplinary project aims at extending the definition and the range of applicability of distance geometry, with a particular attention to its discretization. As it is already possible to remark from recent publications in the scientific literature, the distance geometry can nowadays be seen as a classical problem in operational research, with a wide range of potential applications. Among the possible extensions, this project will mainly focus on dynamical problems, motivated by a certain number of novel applications that we have identified. These include interaction motion adaptation, the simulation of crowd behaviors, and the conception of recommender systems that are able to satisfy modern privacy regulations. The classical application of the distance geometry arising in the biological field will also be considered in this project. The necessity of a strong computational power for the mentioned applications motivates the need of implementing our algorithms in environments capable of exploiting the resources in GPU cards.

Partners are: Inria, Unviersité de Rennes 2, INSA Rennes, Université d'Avignon, CNRS.

9.2. European Initiatives

9.2.1. FP7 & H2020 Projects

9.2.1.1. ANTAREX

Participants: Erven Rohou, Imane Lasri.

Title: Auto-Tuning and Adaptivity appRoach for Energy efficient exascale HPC Systems

Programm: H2020

Duration: September 2015 - September 2018

Coordinator: Politecnico di Milano, Italy (POLIMI)

Partners:

- Consorzio Interuniversitario Cineca (Italy)
- Dompé Farmaceutici Spa (Italy)
- Eidgenoessische Technische Hochschule Zürich (Switzerland)
- Vysoka Skola Banska - Technicka Univerzita Ostrava (Czech Republic)
- Politecnico di Milano (Italy)
- Sygic As (Slovakia)
- Universidade do Porto (Portugal)

Inria contact: Erven Rohou

Energy-efficient heterogeneous supercomputing architectures need to be coupled with a radically new software stack capable of exploiting the benefits offered by the heterogeneity at all the different levels (supercomputer, job, node) to meet the scalability and energy efficiency required by Exascale supercomputers. ANTAREX will solve these challenging problems by proposing a disruptive holistic approach spanning all the decision layers composing the supercomputer software stack and exploiting effectively the full system capabilities (including heterogeneity and energy management). The main goal of the ANTAREX project is to provide a breakthrough approach to express application self-adaptivity at design-time and to runtime manage and autotune applications for green and heterogenous High Performance Computing (HPC) systems up to the Exascale level.

9.2.1.2. Eurolab-4-HPC

Participant: André Seznec.

Title: EuroLab-4-HPC: Foundations of a European Research Center of Excellence in High Performance Computing Systems

Programm: H2020

Duration: September 2015 - September 2017

Coordinator: CHALMERS TEKNISKA HOEGSKOLA AB

Partners:

Barcelona Supercomputing Center - Centro Nacional de Supercomputacion (Spain)

Chalmers Tekniska Hoegskola (Sweden)

École Polytechnique Federale de Lausanne (Switzerland)

Foundation for Research and Technology Hellas (Greece)

Universität Stuttgart (Germany)

Rheinisch-Westfaelische Technische Hochschule Aachen (Germany)

Technion - Israel Institute of Technology (Israel)

Universitaet Augsburg (Germany)

The University of Edinburgh (United Kingdom)

Universiteit Gent (Belgium)

The University of Manchester (United Kingdom)

Inria contact: Albert Cohen (Inria Paris)

Europe has built momentum in becoming a leader in large parts of the HPC ecosystem. It has brought together technical and business stakeholders from application developers via system software to exascale systems. Despite such gains, excellence in high performance computing systems is often fragmented and opportunities for synergy missed. To compete internationally, Europe must bring together the best research groups to tackle the longterm challenges for HPC. These typically cut across layers, e.g., performance, energy efficiency and dependability, so excellence in research must target all the layers in the system stack. The EuroLab-4-HPC project's bold overall goal is to build connected and sustainable leadership in high-performance computing systems by bringing together the different and leading performance oriented communities in Europe, working across all layers of the system stack and, at the same time, fueling new industries in HPC.

9.2.1.3. DAL

Participants: Pierre Michaud, Sylvain Collange, Erven Rohou, André Seznec, Arthur Perais, Sajith Kalathingal, Andrea Mondelli, Aswinkumar Sridharan.

Title: DAL: Defying Amdahl's Law

Program: FP7

Type: ERC

Duration: April 2011 - March 2016

Coordinator: Inria

Inria contact: André Seznec

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 1000's 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 thread. We envision that, around 2020, the processor chips will feature a few complex cores and many (may be 1000's) simpler, more silicon and power effective cores. In the DAL research project, we will 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 will focus on enhancing single process performance. On the microarchitecture side, we will explore both a radically new approach, the sequential accelerator, and more conventional processor architectures. We will also study how to exploit heterogeneous multicore architectures to enhance sequential thread performance.

9.2.1.4. ARGO

Participants: Isabelle Puaut, Damien Hardy.

Title: Argo: WCET-Aware Parallelization of Model-Based Applications for Heterogeneous Parallel Systems

Program: H2020

Type: RIA

Duration: Jan 2016 - Dec 2018

Coordinator: Karlsruher Institut fuer Technologie (KIT)

Université Rennes I contact: Steven Derrien

Partners:

Karlsruher Institut fuer Technologie (KIT)

SCILAB enterprises SAS

Recore Systems BV

Université de Rennes 1

Technologiko Ekpaideftiko Idryma (TEI) Dytikis Elladas

Absint GmbH

Deutsches Zentrum fuer Luft - und Raumfahrt EV

Fraunhofer

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

Argo also involves Steven Derrien, Angeliki Kritikakou, and Imen Fassi from the CAIRN team.

9.2.2. Collaborations in European Programs, Except FP7 & H2020

9.2.2.1. COST Action TACLe - Timing Analysis on Code-Level 10-2012/09-2016

Participants: Damien Hardy, Isabelle Puaut, Benjamin Rouxel.

Embedded systems increasingly permeate our daily lives. Many of those systems are business- or safety-critical, with strict timing requirements. Code-level timing analysis (used to analyze software running on some given hardware w.r.t. its timing properties) is an indispensable technique for ascertaining whether or not these requirements are met. However, recent developments in hardware, especially multi-core processors, and in software organization render analysis increasingly more difficult, thus challenging the evolution of timing analysis techniques.

New principles for building "timing-composable" embedded systems are needed in order to make timing analysis tractable in the future. This requires improved contacts within the timing analysis community, as well as with related communities dealing with other forms of analysis such as model-checking and type-inference, and with computer architectures and compilers. The goal of this COST Action is to gather these forces in order to develop industrial-strength code-level timing analysis techniques for future-generation embedded systems, through several working groups:

- WG1 Timing models for multi-cores and timing composability
- WG2 Tooling aspects
- WG3 Early-stage timing analysis
- WG4 Resources other than time

Isabelle Puaut is in the management committee of the COST Action TACLe - Timing Analysis on Code-Level (http://www.tacle.eu). She is responsible of Short Term Scientific Missions (STSM) within TACLe.

9.2.3. Collaborations with Major European Organizations

9.2.3.1. HiPEAC4 NoE

Participants: Pierre Michaud, Erven Rohou, André Seznec.

P. Michaud, A. Seznec and E. Rohou are members of the European Network of Excellence HiPEAC4.

HiPEAC4 addresses the design and implementation of high-performance commodity computing devices in the 10+ year horizon, covering both the processor design, the optimizing compiler infrastructure, and the evaluation of upcoming applications made possible by the increased computing power of future devices.

9.3. International Initiatives

9.3.1. PHC IMHOTEP

Participant: Erven Rohou.

Title: Thoth – An Automatic Dynamic Binary Parallelisation System

International Partner (Institution - Laboratory - Researcher):

Egypt-Japan University of Science and Technology - Prof. Ahmed ElMahdy.

Dates: 2016–2017

With the current global trend towards utilizing cloud computing and smart devices, executing the same application across becomes a necessity. Moreover, parallelism is now abundant with various forms that include thread- and data-parallel execution models. Such diversity in ISA and explicit parallelism makes software development cost prohibitive, especially for natively optimized binaries. This project leverages dynamic binary translation technology to provide for exploiting the underlying parallel resources without the need of having the source code of the application. In particular the project integrates low overhead dynamic profiling, novel OSR parallel de-optimization and a retargetable parallelization modules to allow for dynamic parallelization of binaries.

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

9.3.2.1. PROSPIEL

Participant: Sylvain Collange.

Title: Profiling and specialization for locality

International Partner (Institution - Laboratory - Researcher):

Universidade Federal de Minas Gerais (Brazil) - DCC - Fernando Magno Quintão Pereira

Start year: 2015

See also: https://team.inria.fr/pacap/prospiel/

The PROSPIEL project aims at optimizing parallel applications for high performance on new throughput-oriented architectures: GPUs and many-core processors. Traditionally, code optimization is driven by a program analysis performed either statically at compile-time, or dynamically at run-time. Static program analysis is fully reliable but often over-conservative. Dynamic analysis provides more accurate data, but faces strong execution time constraints and does not provide any guarantee. By combining profiling-guided specialization of parallel programs with runtime checks for correctness, PROSPIEL seeks to capture the advantages of both static analysis and dynamic analysis. The project relies on the polytope model, a mathematical representation for parallel loops, as a theoretical foundation. It focuses on analyzing and optimizing performance aspects that become increasingly critical on modern parallel computer architectures: locality and regularity.

9.3.3. Inria International Partners

9.3.3.1. Informal International Partners

The PACAP project-team has informal collaborations (visits, common publications) with University of Wisconsin at Madison (Pr Wood), University of Toronto (Pr Moshovos), University of Ghent (Dr Eyerman), University of Uppsala (Pr Hagersten), University of Cyprus (Pr Sazeides), the Egyptian-Japanese University of Science and Technology (Pr Ahmed El-Mahdy), Intel Haifa (Dr Zaks, Eng Nuzman), Barcelona Supercomputing Center (Dr Cazorla, Dr Abella), ISEP Porto (Dr Nelissen, Dr Nélis).

9.4. International Research Visitors

9.4.1. Visits of International Scientists

9.4.1.1. Internships

Rubens Emilio Alves Moreira, student at Universidade Federal de Minas Gerais, visited from Feb 2016 to May 2016 within the context of the PROSPIEL associated team.

Stefano Cherubin, PhD student at Politecnico di Milano for one month in Oct 2016, within the context of the ANTAREX H2020 project.

Anita Tino, PhD student at Ryerson University, visited from Oct 2016 within the context of a MITACS grant.

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events Organisation

10.1.1.1. Member of the Organizing Committees

A. Seznec is member of the ACM/IEEE PACT conference steering committee.

A. Seznec is member of the ACM/IEEE ISCA symposium steering committee.

10.1.2. Scientific Events Selection

10.1.2.1. Chair of Conference Program Committees

Isabelle Puaut is Program Chair of the 2017 IEEE Real-Time Systems Symposium (RTSS).

A. Seznec was PC chair of the 2016 ACM/IEEE ISCA symposium.

10.1.2.2. Member of the Conference Program Committees

Isabelle Puaut is member of the program committees of the Euromicro Conference on Real Time Systems (ECRTS) 2016 and 2017, the IEEE Real-Time Systems Symposium (RTSS) 2016, the IEEE Real-Time and Embedded Technology and Applications Symposium (RTAS) 2017 and the WCET workshop 2016.

A. Seznec is a member of IEEE Micro 2017 Top Picks selection committee.

A. Seznec was a member of the SAMOS 2016 conference program committee.

Damien Hardy was a member of RTNS 2016 and WCET 2016 program committees.

Pierre Michaud was a member of the program committees of the HPCA 2017 conference and of the 5th JILP Workshop on Computer Architecture Competitions (JWAC-5).

Sylvain Collange was PC member of ISCA 2016 and of Compas'2016.

10.1.3. Journal

10.1.3.1. Member of the Editorial Boards

Isabelle Puaut is Associate Editor for IEEE Transactions on Computers (IEEE TC).

A. Seznec is a member of the editorial boards of IEEE Micro and ACM Transactions on Architecture and Compiler Optimization.

10.1.4. Invited Talks

Damien Hardy was invited to give a tutorial on Heptane at Tutor16 (1st Tutorial on Tools for Real-Time Systems) in conjunction with the Cyber-Physical Systems week 2016.

Damien Hardy was invited from August 31st to September 2nd at the Barcelona Supercomputing Center group. He presented an invited talk.

A. Seznec presented invited talks at Intel Bangalore (compressed caches, branch prediction, value prediction) in Sept. 2016.

A . Seznec presented the PACAP work on compressed caches at the ARM research summit in Sept. 2016.

A. Seznec presented the PACAP work on compressed caches and register equality prediction at the Intel low latency ISRA workshop in Dec. 2016.

10.1.5. Scientific Expertise

Erven Rohou was an expert for the ANR review process.

10.1.6. Research Administration

Isabelle Puaut is responsible of Short Term Scientific Missions (STSM) withing the European COST action Tacle (Timing Analysis at Code LEvel) (http://www.tacle.eu).

Isabelle Puaut is member of the steering committee of RTNS (Real-Time Networks and Systems).

Isabelle Puaut is member of the steering committee of the Worst Case Execution Time (WCET) workshop, held in conjunction with the Euromicro Conference on Real Time Systems (ECRTS).

Isabelle Puaut is member of the scientific council of University of Rennes 1.

Isabelle Puaut is member of the administration council of the computer science and electrical engineering department of University of Rennes 1.

A. Seznec is an elected member of the Administration Council of Inria.

Erven Rohou is a member of the Inria CDT (Commission du Développement Technologique).

As "correspondant scientifique des relations internationales" for Inria Rennes Bretagne Atlantique, Erven Rohou is a member of the Inria COST GTRI (Groupe de Travail "Relations Internationales" du Comité d'Orientation Scientifique et Technologique).

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

Master: Isabelle Puaut, Operating systems, 1st year of master, total of 110 hours

Master: Isabelle Puaut, Damien Hardy, Real-time systems, 1st year of master, total of 58 hours

Master: Isabelle Puaut, Erven Rohou, Writing of scientific publications, 2nd year of master and PhD students, total of 24 hours

Licence: Damien Hardy, Real-time systems, L3 Université de Rennes I, total of 60 hours

Master: Damien Hardy, Operating systems, M2 Université de Rennes I, total of 60 hours

Master: Damien Hardy, Operating systems, M1 Université de Rennes I, total of 60 hours

Master: S. Collange, Programmation parallèle, 22 hours, M1, Université de Rennes I, France

10.2.2. Supervision

PhD: Sajith Kalathingal, "Transforming TLP into DLP with the Dynamic Inter-Thread Vectorization Architecture", Université Rennes 1, Dec 2016, co-advisors S. Collange and A. Seznec

PhD: Aswinkumar Sridharan, "Adaptive and Intelligent Memory Systems", Université Rennes 1, Dec. 2016, advisor A. Seznec

PhD: Arjun Suresh, "Intercepting Functions for Memoization", Université Rennes 1, May 2016, co-advisors E. Rohou and A. Seznec

PhD in progress, Viet Anh Nguyen, Worst-Case Execution Time (WCET) Estimation for Many-core Architectures, started in january 2015. Supervised by Isabelle Puaut and Damien Hardy.

PhD in progress, Benjamin Rouxel, Code optimizations for WCET calculation on many-core platforms, started in october 2015. Supervised by Isabelle Puaut and Steven Derrien from the CAIRN group.

PhD in progress: Nabil Hallou, Université Rennes 1, Feb 2013, co-advisors E. Rohou and P. Clauss (EPI Camus Inria Strasbourg)

PhD in progress: Andrea Mondelli, Université Rennes 1, Oct 2013, co-advisors P. Michaud and A. Seznec

PhD in progress: Rabab Bouziane, Université Rennes 1, Nov 2015, advisor E. Rohou and Abdoulaye Gamatié (LIRMM, Montpellier)

PhD in progress: Arif Ali Ana-Pparakkal, Université Rennes 1, Feb 2015, advisor E. Rohou

PhD in progress: Simon Rokicki, Université Rennes 1, Sep 2015, co-advisors E. Rohou and Steven Derrien (CAIRN)

PhD in progress: Kleovoulos, Kalitzidis, "Ultrawide Issue Superscalar Processors", Université Rennes 1, Dec. 2016, advisor A. Seznec

10.2.3. Juries

Isabelle Puaut was a member of the following committees:

- PhD: Pierre Wilke, Formally Verified Compilation of Low-Level C code, Université de Rennes 1, Nov 2016
- PhD: Guillaume Phavorin, Hard Real-Time Scheduling subjected to Cache-Related Preemption Delays, Université de Poitiers, Sep 2016 (rapporteur)
- PhD: Vincent Mussot, Automates d'annotation de flot pour l'expression et l'intégration de propriétés dans l'analyse de WCET, Université Paul Sabatier, Toulouse, Dec 2016 (rapporteur)
- HDR: Mathieu Jan, Contributions au paradigme par cadencement temporel (TT) et à l'embarquabilité des systèmes temps réel, Université Paris Sud, Dec 2016 (rapporteur)

Erven Rohou was a member of the following committees:

- PhD: Juan Manuel Martinez Caamano, Strasbourg
- PhD: Lénaïc Bagnères, Orsay
- PhD: Michele Scandale, Politecnico di Milano, Milan, Italy
- PhD: Amir Ashouri, Politecnico di Milano, Milan, Italy
- PhD: Sébastien Martinez, Télécom Bretagne, Brest
- PhD: Reem ElKhouly, Egypt-Japan University of Science and Technology, Alexandrie, Egypt

10.2.3.1. Assistant professor hiring committees

Isabelle Puaut: University of Toulouse (computer architecture and real-time systems)

Isabelle Puaut: University of Chalmers, Sweden (real-time systems)

10.2.3.2. Professor hiring committee:

Isabelle Puaut: UBO (Université de Bretagne Occidentale) - real-time systems

10.3. Popularization

Erven Rohou discussed the research axes of the team in the “émergences” newsletter http://emergences.inria.fr/2016/newsletter_n43/L42-PACAP.

Nicolas Kiss, Damien Hardy and Erven Rohou presented a poster at the “Rencontres inter-UMRs-DGA”, of the “Pôle d’excellence Cyber”.

Erven Rohou and Isabelle Puaut presented a poster (with ANR W-SEPT colleagues) at “Les rencontres du numérique de l’ANR”.

11. Bibliography

Major publications by the team in recent years

[1] F. BODIN, T. KISUKI, P. M. W. KNIJNENBURG, M. F. P. O’BOYLE, E. ROHOU.*Iterative Compilation in a Non-Linear Optimisation Space*, in "Workshop on Profile and Feedback-Directed Compilation (FDO-1), in conjunction with PACT ’98", October 1998.

[2] A. COHEN, E. ROHOU.*Processor Virtualization and Split Compilation for Heterogeneous Multicore Embedded Systems*, in "DAC", June 2010, p. 102–107.

[3] N. HALLOU, E. ROHOU, P. CLAUSS, A. KETTERLIN.*Dynamic Re-Vectorization of Binary Code*, in "SAMOS", July 2015, https://hal.inria.fr/hal-01155207.

[4] D. HARDY, I. PUAUT.*Static probabilistic Worst Case Execution Time Estimation for architectures with Faulty Instruction Caches*, in "21st International Conference on Real-Time Networks and Systems", Sophia Antipolis, France, October 2013 [*DOI :* 10.1145/2516821.2516842], https://hal.inria.fr/hal-00862604.

[5] D. HARDY, I. SIDERIS, A. SAIDI, Y. SAZEIDES.*EETCO: A tool to estimate and explore the implications of datacenter design choices on the tco and the environmental impact*, in "Workshop on Energy-efficient Computing for a Sustainable World in conjunction with the 44th Annual IEEE/ACM International Symposium on Microarchitecture (Micro-44)", 2011.

[6] M.-K. LEE, P. MICHAUD, J. S. SIM, D. NYANG.*A simple proof of optimality for the MIN cache replacement policy*, in "Information Processing Letters", September 2015, 3 [*DOI :* 10.1016/J.IPL.2015.09.004], https://hal.inria.fr/hal-01199424.

[7] P. MICHAUD.*A Best-Offset Prefetcher* ***Champion***, in "2nd Data Prefetching Championship", Portland, OR, USA, June 2015, https://hal.inria.fr/hal-01165600.

[8] P. MICHAUD, A. MONDELLI, A. SEZNEC.*Revisiting Clustered Microarchitecture for Future Superscalar Cores: A Case for Wide Issue Clusters*, in "ACM Transactions on Architecture and Code Optimization (TACO) ", August 2015, vol. 13, n° 3, 22 [*DOI :* 10.1145/2800787], https://hal.inria.fr/hal-01193178.

[9] P. MICHAUD, A. SEZNEC.*Pushing the branch predictability limits with the multi-poTAGE+SC predictor :* ***Champion in the unlimited category***, in "4th JILP Workshop on Computer Architecture Competitions (JWAC-4): Championship Branch Prediction (CBP-4)", Minneapolis, United States, June 2014, https://hal.archives-ouvertes.fr/hal-01087719.

[10] A. PERAIS.*Increasing the performance of superscalar processors through value prediction*, Université Rennes 1, September 2015, https://tel.archives-ouvertes.fr/tel-01282474.

[11] A. PERAIS, A. SEZNEC.*EOLE: Paving the Way for an Effective Implementation of Value Prediction*, in "International Symposium on Computer Architecture", Minneapolis, MN, United States, ACM/IEEE, June 2014, vol. 42, p. 481 - 492 [*DOI :* 10.1109/ISCA.2014.6853205], https://hal.inria.fr/hal-01088130.

[12] A. PERAIS, A. SEZNEC.*Practical data value speculation for future high-end processors*, in "International Symposium on High Performance Computer Architecture", Orlando, FL, United States, IEEE, February 2014, p. 428 - 439 [*DOI :* 10.1109/HPCA.2014.6835952], https://hal.inria.fr/hal-01088116.

[13] A. PERAIS, A. SEZNEC.*EOLE: Toward a Practical Implementation of Value Prediction*, in "IEEE Micro", June 2015, vol. 35, n° 3, p. 114 - 124 [*DOI :* 10.1109/MM.2015.45], https://hal.inria.fr/hal-01193287.

[14] E. RIOU, E. ROHOU, P. CLAUSS, N. HALLOU, A. KETTERLIN.*PADRONE: a Platform for Online Profiling, Analysis, and Optimization*, in "Dynamic Compilation Everywhere", Vienna, Austria, January 2014.

[15] S. SARDASHTI, A. SEZNEC, D. A. WOOD.*Skewed Compressed Caches*, in "47th Annual IEEE/ACM International Symposium on Microarchitecture, 2014", Minneapolis, United States, December 2014, https://hal.inria.fr/hal-01088050.

[16] A. SEMBRANT, T. CARLSON, E. HAGERSTEN, D. BLACK-SHAFFER, A. PERAIS, A. SEZNEC, P. MICHAUD.*Long Term Parking (LTP): Criticality-aware Resource Allocation in OOO Processors*, in "International Symposium on Microarchitecture, Micro 2015", Honolulu, United States, Proceeding of the International Symposium on Microarchitecture, Micro 2015, ACM, December 2015, https://hal.inria.fr/hal-01225019.

[17] A. SEZNEC, P. MICHAUD.*A case for (partially)-tagged geometric history length predictors*, in "Journal of Instruction Level Parallelism", April 2006, http://www.jilp.org/vol8.

[18] A. SEZNEC, J. SAN MIGUEL, J. ALBERICIO.*The Inner Most Loop Iteration counter: a new dimension in branch history* , in "48th International Symposium On Microarchitecture", Honolulu, United States, ACM, December 2015, 11, https://hal.inria.fr/hal-01208347.

[19] A. SEZNEC.*TAGE-SC-L Branch Predictors: **Champion in 32Kbits and 256 Kbits category***, in "JILP - Championship Branch Prediction", Minneapolis, United States, June 2014, https://hal.inria.fr/hal-01086920.

[20] A. SURESH, B. NARASIMHA SWAMY, E. ROHOU, A. SEZNEC.*Intercepting Functions for Memoization: A Case Study Using Transcendental Functions*, in "ACM Transactions on Architecture and Code Optimization (TACO)", July 2015, vol. 12, n^{o} 2, 23 [*DOI :* 10.1145/2751559], https://hal.inria.fr/hal-01178085.

[21] D. D. C. TEIXEIRA, S. COLLANGE, F. M. Q. PEREIRA.*Fusion of calling sites*, in "International Symposium on Computer Architecture and High-Performance Computing (SBAC-PAD)", Florianópolis, Santa Catarina, Brazil, October 2015 [*DOI :* 10.1109/SBAC-PAD.2015.16], https://hal.archives-ouvertes.fr/hal-01410221.

Publications of the year

Doctoral Dissertations and Habilitation Theses

[22] S. KALATHINGAL.*Transforming TLP into DLP with the Dynamic Inter-Thread Vectorization Architecture*, Université Rennes 1, December 2016, https://tel.archives-ouvertes.fr/tel-01426915.

[23] A. SRIDHARAN.*Adaptive and Intelligent Memory Systems*, Inria Rennes - Bretagne Atlantique and University of Rennes 1, France, December 2016, https://hal.inria.fr/tel-01442465.

[24] A. SURESH.*Intercepting Functions for Memoization*, Université de Rennes 1, May 2016, https://hal.inria.fr/tel-01410539.

Articles in International Peer-Reviewed Journal

[25] P. MICHAUD.*Some mathematical facts about optimal cache replacement*, in "ACM Transactions on Architecture and Code Optimization (TACO) ", December 2016, vol. 13, n^{o} 4 [*DOI :* 10.1145/3017992], https://hal.inria.fr/hal-01411156.

[26] B. PANDA.*SPAC: A Synergistic Prefetcher Aggressiveness Controller for Multi-core Systems*, in "IEEE Transactions on Computers", 2016 [*DOI :* 10.1109/TC.2016.2547392], https://hal.inria.fr/hal-01307538.

[27] A. PERAIS, A. SEZNEC.*EOLE: Combining Static and Dynamic Scheduling through Value Prediction to Reduce Complexity and Increase Performance*, in "TOCS - ACM Transactions on Computer Systems", February 2016, 34, https://hal.inria.fr/hal-01259139.

[28] A. PERAIS, A. SEZNEC.*Storage-Free Memory Dependency Prediction*, in "IEEE Computer Architecture Letters", November 2016, p. 1 - 4 [*DOI :* 10.1109/LCA.2016.2628379], https://hal.inria.fr/hal-01396985.

[29] S. SARDASHTI, A. SEZNEC, D. A. WOOD.*Yet Another Compressed Cache: a Low Cost Yet Effective Compressed Cache*, in "ACM Transactions on Architecture and Code Optimization", September 2016, 25, https://hal.inria.fr/hal-01354248.

[30] A. SEZNEC, J. SAN MIGUEL, J. ALBERICIO.*Practical Multidimensional Branch Prediction*, in "IEEE Micro", 2016 [*DOI :* 10.1109/MM.2016.33], https://hal.inria.fr/hal-01330510.

Invited Conferences

[31] R. IAKYMCHUK, D. DEFOUR, S. COLLANGE, S. GRAILLAT.*Reproducible and Accurate Algorithms for Numerical Linear Algebra*, in "PP: Parallel Processing for Scientific Computing", Paris, France, SIAM, April 2016, https://hal-lirmm.ccsd.cnrs.fr/lirmm-01268048.

[32] C. SILVANO, G. AGOSTA, S. CHERUBIN, D. GADIOLI, G. PALERMO, A. BARTOLINI, L. BENINI, J. MARTINOVIČ, M. PALKOVIČ, K. SLANINOVÁ, J. BISPO, J. M. P. CARDOSO, R. ABREU, P. PINTO, C. CAVAZZONI, N. SANNA, A. R. BECCARI, R. CMAR, E. ROHOU.*The ANTAREX Approach to Autotuning and Adaptivity for Energy Efficient HPC Systems*, in "ACM International Conference on Computing Frontiers 2016", Como, Italy, May 2016 [*DOI :* 10.1145/2903150.2903470], https://hal.inria.fr/hal-01341826.

International Conferences with Proceedings

[33] A. BONENFANT, F. CARRIER, H. CASSÉ, P. CUENOT, D. CLARAZ, N. HALBWACHS, H. LI, C. MAIZA, M. DE MICHIEL, V. MUSSOT, C. PARENT-VIGOUROUX, I. PUAUT, P. RAYMOND, E. ROHOU, P. SOTIN.*When the worst-case execution time estimation gains from the application semantics*, in "8th European Congress on Embedded Real-Time Software and Systems", Toulouse, France, January 2016, https://hal.inria.fr/hal-01235781.

[34] S. COLLANGE, M. JOLDES, J.-M. MULLER, V. POPESCU.*Parallel floating-point expansions for extended-precision GPU computations*, in "The 27th Annual IEEE International Conference on Application-specific Systems, Architectures and Processors (ASAP)", London, United Kingdom, July 2016, https://hal.archives-ouvertes.fr/hal-01298206.

[35] *Best Paper*
D. HARDY, I. PUAUT, Y. SAZEIDES.*Probabilistic WCET estimation in presence of hardware for mitigating the impact of permanent faults*, in "Design, Automation and Test in Europe", Dresden, Germany, March 2016, https://hal.inria.fr/hal-01259493.

[36] *Best Paper*
S. KALATHINGAL, S. COLLANGE, B. NARASIMHA SWAMY, A. SEZNEC.*Dynamic Inter-Thread Vectorization Architecture: extracting DLP from TLP*, in "International Symposium on Computer Architecture and High-Performance Computing (SBAC-PAD)", Los Angeles, United States, October 2016, https://hal.inria.fr/hal-01356202.

[37] P. MICHAUD.*Best-Offset Hardware Prefetching*, in "International Symposium on High-Performance Computer Architecture", Barcelona, Spain, March 2016 [*DOI :* 10.1109/HPCA.2016.7446087], https://hal.inria.fr/hal-01254863.

[38] R. E. A. MOREIRA, S. COLLANGE, F. M. Q. PEREIRA.*Function Call Re-Vectorization*, in "ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming (PPoPP)", Austin, Texas, United States, February 2017, https://hal.archives-ouvertes.fr/hal-01410186.

[39] B. PANDA, A. SEZNEC.*Dictionary Sharing: An Efficient Cache Compression Scheme for Compressed Caches*, in "49th Annual IEEE/ACM International Symposium on Microarchitecture, 2016", Taipei, Taiwan, IEEE/ACM, October 2016, https://hal.archives-ouvertes.fr/hal-01354246.

[40] A. PERAIS, F. A. ENDO, A. SEZNEC.*Register Sharing for Equality Prediction*, in "International Symposium on Microarchitecture", Taipei, Taiwan, October 2016, https://hal.inria.fr/hal-01354267.

[41] A. PERAIS, A. SEZNEC.*Cost Effective Physical Register Sharing*, in "International Symposium on High Performance Computer Architecture", Barcelona, Spain, IEEE, March 2016, vol. 22 [*DOI :* 10.1109/HPCA.2016.7446105], https://hal.inria.fr/hal-01259137.

[42] P.-Y. PÉNEAU, R. BOUZIANE, A. GAMATIÉ, E. ROHOU, F. BRUGUIER, G. SASSATELLI, L. TORRES, S. SENNI.*Loop Optimization in Presence of STT-MRAM Caches: a Study of Performance-Energy Tradeoffs*, in "26th International Workshop on Power and Timing Modeling, Optimization and Simulation", Bremen, Germany, Proceedings of the 26th International Workshop on Power and Timing Modeling, Optimization and Simulation, September 2016, 8, https://hal.inria.fr/hal-01347354.

[43] S. A. RASHID, G. NELISSEN, D. HARDY, B. AKESSON, I. PUAUT, E. TOVAR.*Cache-Persistence-Aware Response-Time Analysis for Fixed-Priority Preemptive Systems*, in "28th Euromicro Conference on Real-Time Systems (ECRTS)", Toulouse, France, IEEE, July 2016 [*DOI :* 10.1109/ECRTS.2016.25], https://hal.inria.fr/hal-01393220.

[44] S. ROKICKI, E. ROHOU, S. DERRIEN.*Hardware-Accelerated Dynamic Binary Translation*, in "IEEE/ACM Design, Automation & Test in Europe Conference & Exhibition (DATE)", Lausanne, Switzerland, March 2017, https://hal.inria.fr/hal-01423639.

[45] *Best Paper*
A. SEZNEC.*Exploring branch predictability limits with the MTAGE+SC predictor* *, in "5th JILP Workshop on Computer Architecture Competitions (JWAC-5): Championship Branch Prediction (CBP-5)", Seoul, South Korea, June 2016, 4, https://hal.inria.fr/hal-01354251.

[46] *Best Paper*
A. SEZNEC.*TAGE-SC-L Branch Predictors Again*, in "5th JILP Workshop on Computer Architecture Competitions (JWAC-5): Championship Branch Prediction (CBP-5)", Seoul, South Korea, June 2016, https://hal.inria.fr/hal-01354253.

[47] C. SILVANO, G. AGOSTA, A. BARTOLINI, A. R. BECCARI, L. BENINI, J. BISPO, R. CMAR, J. M. P. CARDOSO, C. CAVAZZONI, J. MARTINOVIČ, G. PALERMO, M. PALKOVIČ, P. PINTO, E. ROHOU, N. SANNA, K. SLANINOVÁ.*AutoTuning and Adaptivity appRoach for Energy efficient eXascale HPC systems: the ANTAREX Approach*, in "Design, Automation, and Test in Europe", Dresden, Germany, Design, Automation, and Test in Europe, March 2016, https://hal.inria.fr/hal-01235741.

[48] *Best Paper*
A. SRIDHARAN, A. SEZNEC.*Discrete Cache Insertion Policies for Shared Last Level Cache Management on Large Multicores*, in "30th IEEE International Parallel & Distributed Processing Symposium", Chicago, United States, May 2016, https://hal.inria.fr/hal-01259626.

[49] A. SURESH, E. ROHOU, A. SEZNEC.*Compile-Time Function Memoization*, in "26th International Conference on Compiler Construction", Austin, United States, February 2017, https://hal.inria.fr/hal-01423811.

National Conferences with Proceeding

[50] S. COLLANGE.*A Synthesizable General-Purpose SIMT Processor*, in "Conférence d'informatique en Parallélisme, Architecture et Système (Compas)", Lorient, France, July 2016, https://hal.inria.fr/hal-01345070.

[51] S. ROKICKI, E. ROHOU, S. DERRIEN.*Hybrid-JIT : Compilateur JIT Matériel/Logiciel pour les Processeurs VLIW Embarqués*, in "Confe´rence d'informatique en Paralle´lisme, Architecture et Syste`me (Compas)", Lorient, France, July 2016, https://hal.archives-ouvertes.fr/hal-01345306.

Research Reports

[52] S. COLLANGE.*Simty: a Synthesizable General-Purpose SIMT Processor*, Inria Rennes Bretagne Atlantique, August 2016, n^{o} RR-8944, https://hal.inria.fr/hal-01351689.

[53] S. SARDASHTI, A. SEZNEC, D. A. WOOD.*Yet Another Compressed Cache: a Low Cost Yet Effective Compressed Cache*, Inria, February 2016, n^{o} RR-8853, 23, https://hal.inria.fr/hal-01270792.

References in notes

[54] L. A. BELADY.*A study of replacement algorithms for a virtual-storage computer*, in "IBM Systems Journal", 1966, vol. 5, n^{o} 2, p. 78-101.

[55] M. HATABA, A. EL-MAHDY, E. ROHOU.*OJIT: A Novel Obfuscation Approach Using Standard Just-In-Time Compiler Transformations*, in "International Workshop on Dynamic Compilation Everywhere", January 2015.

[56] R. KUMAR, D. M. TULLSEN, N. P. JOUPPI, P. RANGANATHAN.*Heterogeneous chip multiprocessors*, in "IEEE Computer", nov. 2005, vol. 38, n^{o} 11, p. 32–38.

[57] R. L. MATTSON, J. GECSEI, D. R. SLUTZ, I. L. TRAIGER.*Evaluation techniques for storage hierarchies*, in "IBM Systems Journal", 1970, vol. 9, n^{o} 2, p. 78-117.

[58] S. NASSIF, N. MEHTA, Y. CAO.*A resilience roadmap*, in "Design, Automation Test in Europe Conference Exhibition (DATE), 2010", March 2010, p. 1011-1016.

[59] J. NICKOLLS, W. J. DALLY.*The GPU computing era*, in "Micro, IEEE", 2010, vol. 30, n^{o} 2, p. 56–69.

[60] R. OMAR, A. EL-MAHDY, E. ROHOU.*Arbitrary control-flow embedding into multiple threads for obfuscation: a preliminary complexity and performance analysis*, in "Proceedings of the 2nd international workshop on Security in cloud computing", ACM, 2014, p. 51–58.

[61] S. SARDASHTI, D. A. WOOD.*Decoupled compressed cache: exploiting spatial locality for energy-optimized compressed caching*, in "The 46th Annual IEEE/ACM International Symposium on Microarchitecture, MICRO-46, Davis, CA, USA, December 7-11, 2013", 2013, p. 62–73, http://doi.acm.org/10.1145/2540708.2540715.

[62] A. SEZNEC, N. SENDRIER.*HAVEGE: A user-level software heuristic for generating empirically strong random numbers*, in "ACM Transactions on Modeling and Computer Simulation (TOMACS)", 2003, vol. 13, n^{o} 4, p. 334–346.

Project-Team PANAMA

Parcimonie et Nouveaux Algorithmes pour le Signal et la Modélisation Audio

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

IN PARTNERSHIP WITH:
CNRS

Université Rennes 1

RESEARCH CENTER
Rennes - Bretagne-Atlantique

THEME
Language, Speech and Audio

Table of contents

Project-Team PANAMA

Creation of the Project-Team: 2013 January 01

Keywords:

Computer Science and Digital Science:

1.2.6. - Sensor networks
3.1.1. - Modeling, representation
3.3.3. - Big data analysis
3.4.1. - Supervised learning
3.4.2. - Unsupervised learning
3.4.4. - Optimization and learning
3.4.5. - Bayesian methods
3.4.6. - Neural networks
3.4.7. - Kernel methods
3.4.8. - Deep learning
3.5.1. - Analysis of large graphs
5.3.2. - Sparse modeling and image representation
5.7.1. - Sound
5.7.2. - Music
5.7.3. - Speech
5.7.4. - Analysis
5.9.1. - Sampling, acquisition
5.9.2. - Estimation, modeling
5.9.3. - Reconstruction, enhancement
5.9.4. - Signal processing over graphs
5.9.5. - Sparsity-aware processing
5.9.6. - Optimization tools
5.10.2. - Perception
5.11.2. - Home/building control and interaction
6.1.4. - Multiscale modeling
6.2.5. - Numerical Linear Algebra
6.2.6. - Optimization
6.3.1. - Inverse problems
6.3.2. - Data assimilation
7.8. - Information theory
7.9. - Graph theory

Other Research Topics and Application Domains:

1.3. - Neuroscience and cognitive science
2.5.1. - Sensorimotor disabilities
2.6. - Biological and medical imaging
5.6. - Robotic systems
5.8. - Learning and training

6.3.3. - Network Management
8.1.2. - Sensor networks for smart buildings
8.4. - Security and personal assistance
9.1. - Education
9.2.1. - Music, sound
9.2.2. - Cinema, Television
9.2.3. - Video games
9.6. - Reproducibility
9.9.1. - Environmental risks

1. Members

Research Scientists

Rémi Gribonval [Team leader, Senior Researcher, Inria, HDR]
Nancy Bertin [Researcher, CNRS]
Frédéric Bimbot [Senior Researcher, CNRS, HDR]
Antoine Deleforge [Researcher, Inria]

Technical Staff

Ewen Camberlein [CNRS]
Romain Lebarbenchon [CNRS]
Nicolas Bellot [Inria, granted by FP7 ERC PLEASE project]
Adrien Leman [Inria, from Feb 2016, granted by FP7 ERC PLEASE project]

PhD Students

Corentin Guichaoua [Univ. Rennes I]
Luc Le Magoarou [Inria, granted by FP7 ERC PLEASE project]
Nicolas Keriven [Univ. Rennes I, Contrat Doctoral Polytechnicien]
Himalaya Jain [Technicolor, granted by CIFRE Technicolor]
Clément Gaultier [Inria, granted by FP7 ERC PLEASE project and Conseil Régional Bretagne]
Corentin Louboutin [Univ. Rennes I (ENS Rennes)]
Nathan Libermann [Inria, from Oct 2016, granted by ANR DEEPMUS project and Conseil Régional Bretagne]
Cassio Fraga Dantas [Inria, from Oct 2016]
Mohamed Hafsati [IRT b<>com, granted by ANR INVATE project, from Dec 2016]

Post-Doctoral Fellows

Gilles Puy [Inria, until Jan 2016, granted by FP7 ERC PLEASE project]
Nicolas Tremblay [Inria, until Aug 2016]
Yann Traonmilin [Inria, granted by FP7 ERC PLEASE project]
Srdan Kitic [Inria, until Mar 2016, granted by FP7 ERC PLEASE project]

Visiting Scientist

Pierre Vandergheynst [EPFL, Chaire Internationale Inria position, Jul 2015]

Administrative Assistants

Stéphanie Lemaile [Inria]
Armelle Mozziconacci [CNRS, from May 2016]

Others

Saurabh Kataria [intern, from May 2016 until Jul 2016]
Igal Rozenberg [intern, from Apr 2016 until Jul 2016]
Guillaume Versal [intern, from Jun 2016 until Aug 2016]

2. Overall Objectives

2.1. Overall positioning

At the interface between audio modeling and mathematical signal processing, the global objective of PANAMA is to develop mathematically founded and algorithmically efficient techniques to model, acquire and process high-dimensional signals, with a strong emphasis on acoustic data.

Applications fuel the proposed mathematical and statistical frameworks with practical scenarii, and the developed algorithms are extensively tested on targeted applications. PANAMA's methodology relies on a closed loop between theoretical investigations, algorithmic development and empirical studies.

2.2. Scientific foundations

The scientific foundations of PANAMA are focused on sparse representations and probabilistic modeling, and its scientific scope is extended in three major directions:

- The extension of the sparse representation paradigm towards that of "sparse modeling", with the challenge of establishing, strengthening and clarifying connections between sparse representations and machine learning.
- A focus on sophisticated probabilistic models and advanced statistical methods to account for complex dependencies between multi-layered variables (such as in audio-visual streams, musical contents, biomedical data ...).
- The investigation of graph-based representations, processing and transforms, with the goal to describe, model and infer underlying structures within content streams or data sets.

2.3. Applications

The main industrial sectors in relation with the topics of the PANAMA research group are the telecommunication sector, the Internet and multimedia sector, the musical and audiovisual production sector and, marginally, the sector of education and entertainment. Source separation is one of PANAMA's major applicative focus generating increasing industrial transfers. The models, methods and algorithms developed in the team have many potential applications beyond audio processing and modeling – the central theme of the PANAMA project-team – in particular to biomedical signals. Such applications are primarily investigated in partnership with research groups with the relevant expertise (within or outside Inria).

On a regular basis, PANAMA is involved in bilateral or multilateral partnerships, within the framework of consortia, networks, thematic groups, national and European research projects, as well as industrial contracts with various local companies.

3. Research Program

3.1. Axis 1: Sparse Models and Representations

3.1.1. Efficient Sparse Models and Dictionary Design for Large-scale Data

Sparse models are at the core of many research domains where the large amount and high-dimensionality of digital data requires concise data descriptions for efficient information processing. Recent breakthroughs have demonstrated the ability of these models to provide concise descriptions of complex data collections, together with algorithms of provable performance and bounded complexity.

A crucial prerequisite for the success of today's methods is the knowledge of a "dictionary" characterizing how to concisely describe the data of interest. Choosing a dictionary is currently something of an "art", relying on expert knowledge and heuristics.

Pre-chosen dictionaries such as wavelets, curvelets or Gabor dictionaries, are based upon stylized signal models and benefit from fast transform algorithms, but they fail to fully describe the content of natural signals and their variability. They do not address the huge diversity underlying modern data much beyond time series and images: data defined on graphs (social networks, internet routing, brain connectivity), vector valued data (diffusion tensor imaging of the brain), multichannel or multi-stream data (audiovisual streams, surveillance networks, multimodal biomedical monitoring).

The alternative to a pre-chosen dictionary is a trained dictionary learned from signal instances. While such representations exhibit good performance on small-scale problems, they are currently limited to low-dimensional signal processing due to the necessary training data, memory requirements and computational complexity. Whether designed or learned from a training corpus, dictionary-based sparse models and the associated methodology fail to scale up to the volume and resolution of modern digital data, for they intrinsically involve difficult linear inverse problems. To overcome this bottleneck, a new generation of efficient sparse models is needed, beyond dictionaries, encompassing the ability to provide sparse and structured data representations as well as computational efficiency. For example, while dictionaries describe low-dimensional signal models in terms of their "synthesis" using few elementary building blocks called atoms, in "analysis" alternatives the low-dimensional structure of the signal is rather "carved out" by a set of equations satisfied by the signal. Linear as well as nonlinear models can be envisioned.

3.1.2. Compressive Learning

A flagship emerging application of sparsity is the paradigm of compressive sensing, which exploits sparse models at the analog and digital levels for the acquisition, compression and transmission of data using limited resources (fewer/less expensive sensors, limited energy consumption and transmission bandwidth, etc.). Besides sparsity, a key pillar of compressive sensing is the use of random low-dimensional projections. Through compressive sensing, random projections have shown their potential to allow drastic dimension reduction with controlled information loss, provided that the projected signal vector admits a sparse representation in some transformed domain. A related scientific domain, where sparsity has been recognized as a key enabling factor, is Machine Learning, where the overall goal is to design statistically founded principles and efficient algorithms in order to infer general properties of large data collections through the observation of a limited number of representative examples. Marrying sparsity and random low-dimensional projections with machine learning shall allow the development of techniques able to efficiently capture and process the information content of large data collections. The expected outcome is a dramatic increase of the impact of sparse models in machine learning, as well as an integrated framework from the signal level (signals and their acquisition) to the semantic level (information and its manipulation), and applications to data sizes and volumes of collections that cannot be handled by current technologies.

3.2. Axis 2: Robust Acoustic Scene Analysis

3.2.1. Compressive Acquisition and Processing of Acoustic Scenes

Acoustic imaging and scene analysis involve acquiring the information content from acoustic fields with a limited number of acoustic sensors. A full 3D+t field at CD quality and Nyquist spatial sampling represents roughly 10^6 microphones/m^3. Dealing with such high-dimensional data requires to drastically reduce the data flow by positioning appropriate sensors, and selecting from all spatial locations the few spots where acoustic sources are active. The main goal is to develop a theoretical and practical understanding of the conditions under which compressive acoustic sensing is both feasible and robust to inaccurate modeling, noisy measures, and partially failing or uncalibrated sensing devices, in various acoustic sensing scenarii. This requires the development of adequate algorithmic tools, numerical simulations, and experimental data in simple settings where hardware prototypes can be implemented.

3.2.2. Robust Audio Source Separation

Audio signal separation consists in extracting the individual sound of different instruments or speakers that were mixed on a recording. It is now successfully addressed in the academic setting of linear instantaneous mixtures. Yet, real-life recordings, generally associated to reverberant environments, remain an unsolved difficult challenge, especially with many sources and few audio channels. Much of the difficulty comes from the combination of (i) complex source characteristics, (ii) sophisticated underlying mixing model and (iii) adverse recording environments. Moreover, as opposed to the "academic" blind source separation task, most applicative contexts and new interaction paradigms offer a variety of situations in which prior knowledge and adequate interfaces enable the design and the use of informed and/or manually assisted source separation methods.

The former METISS team has developed a generic and flexible probabilistic audio source separation framework that has the ability to combine various acoustic models such as spatial and spectral source models. Building on this existing framework, a first objective of PANAMA is to instantiate and validate specific instances of this framework targeted to real-world industrial applications, such as 5.1 movie re-mastering, interactive music soloist control and outdoor speech enhancement. Extensions of the framework are needed to achieve real-time online processing, and advanced constraints or probabilistic priors for the sources at hand need to be designed, while paying attention to computational scalability issues.

In parallel to these efforts, expected progress in sparse modeling for inverse problems shall bring new approaches to source separation and modeling, as well as to source localization, which is often an important first step in a source separation workflow.

3.2.3. Robust Audio Source Localization

Audio source localization consists in estimating the position of one or several sound sources given the signals received by a microphone array. Knowing the geometry of an audio scene is often a pre-requisite to perform higher-level tasks such as speaker identification and tracking, speech enhancement and recognition or audio source separation. It can be decomposed into two sub-tasks : (i) compute spatial auditory features from raw audio input and (ii) map these features to the desired spatial information. Robustly addressing both these aspects with a limited number of microphones, in the presence of noise, reverberation, multiple and possibly moving sources remains a key challenge in audio signal processing. The first aspect will be tackled by both advanced statistical and acoustical modeling of spatial auditory features. The second one will be addressed by two complementary approaches. *Physics-driven* approaches cast sound source localization as an inverse problem given the known physics of sound propagation within the considered system. *Data-driven* approaches aim at learning the desired feature-to-source-position mapping using real-world or synthetic training datasets adapted to the problem at hand. Combining these approaches should allow a widening of the notion of source localization, considering problems such as the identification of the directivity or diffuseness of the source as well as some of the boundary conditions of the room. A general perspective is to investigate the relations between the physical structure of the source and the particular structures that can be discovered or enforced in the representations and models used for characterization, localization and separation.

3.3. Axis 3: Large-scale Audio Content Processing and Self-organization

3.3.1. Motif Discovery in Audio Data

Facing the ever-growing quantity of multimedia content, the topic of motif discovery and mining has become an emerging trend in multimedia data processing with the ultimate goal of developing weakly supervised paradigms for content-based analysis and indexing. In this context, speech, audio and music content, offers a particularly relevant information stream from which meaningful information can be extracted to create some form of "audio icons" (key-sounds, jingles, recurrent locutions, musical choruses, etc ...) without resorting to comprehensive inventories of expected patterns.

This challenge raises several fundamental questions that will be among our core preoccupations over the next few years. The first question is the deployment of motif discovery on a large scale, a task that requires extending audio motif discovery approaches to incorporate efficient time series pattern matching methods (fingerprinting, similarity search indexing algorithms, stochastic modeling, etc.). The second question is that of the use and interpretation of the motifs discovered. Linking motif discovery and symbolic learning techniques, exploiting motif discovery in machine learning are key research directions to enable the interpretation of recurring motifs.

On the application side, several use cases can be envisioned which will benefit from motif discovery deployed on a large scale. For example, in spoken content, word-like repeating fragments can be used for several spoken document-processing tasks such as language-independent topic segmentation or summarization. Recurring motifs can also be used for audio summarization of audio content. More fundamentally, motif discovery paves the way for a shift from supervised learning approaches for content description to unsupervised paradigms where concepts emerge from the data.

3.3.2. Structure Modeling and Inference in Audio and Musical Contents

Structuring information is a key step for the efficient description and learning of all types of contents, and in particular audio and musical contents. Indeed, structure modeling and inference can be understood as the task of detecting dependencies (and thus establishing relationships) between different fragments, parts or sections of information content.

A stake of structure modeling is to enable more robust descriptions of the properties of the content and better model generalization abilities that can be inferred from a particular content, for instance via cache models, trigger models or more general graphical models designed to render the information gained from structural inference. Moreover, the structure itself can become a robust descriptor of the content, which is likely to be more resistant than surface information to a number of operations such as transmission, transduction, copyright infringement or illegal use.

In this context, information theory concepts need to be investigated to provide criteria and paradigms for detecting and modeling structural properties of audio contents, covering potentially a wide range of application domains in speech content mining, music modeling or audio scene monitoring.

4. Application Domains

4.1. Acoustic Scene Capture

Acoustic fields carry much information about audio sources (musical instruments, speakers, etc.) and their environment (e.g., church acoustics differ much from office room acoustics). A particular challenge is to capture as much information from a complete 3D+t acoustic field associated with an audio scene, using as few sensors as possible. The feasibility of compressive sensing to address this challenge was shown in certain scenarii, and the actual implementation of this framework will potentially impact practical scenarii such as remote surveillance to detect abnormal events, e.g. for health care of the elderly or public transport surveillance.

4.2. Audio Signal Separation in Reverberant Environments

Audio signal separation consists in extracting the individual sound of different instruments or speakers that were mixed on a recording. It is now successfully addressed in the academic setting of linear instantaneous mixtures. Yet, real-life recordings, generally associated to reverberant environments, remain an unsolved difficult challenge, especially with many sources and few audio channels. Much of the difficulty comes from the estimation of the unknown room impulse response associated to a matrix of mixing filters, which can be expressed as a dictionary-learning problem. Solutions to this problem have the potential to impact, for example, the music and game industry, through the development of new digital re-mastering techniques and virtual reality tools, but also surveillance and monitoring applications, where localizing audio sources is important.

4.3. Multimedia Indexing

Audiovisual and multimedia content generate large data streams (audio, video, associated data such as text, etc.). Manipulating large databases of such content requires efficient techniques to: segment the streams into coherent sequences; label them according to words, language, speaker identity, and more generally to the type of content; index them for easy querying and retrieval, etc. As the next generation of online search engines will need to offer content-based means of searching, the need to drastically reduce the computational burden of these tasks is becoming all the more important as we can envision the end of the era of wasteful datacenters that can increase forever their energy consumption. Most of today's techniques to deal with such large audio streams involve extracting features such as Mel Frequency Cepstral Coefficients (MFCC) and learning high-dimensional statistical models such as Gaussian Mixture Models, with several thousand parameters. The exploration of a compressive learning framework is expected to contribute to new techniques to efficiently process such streams and perform segmentation, classification, etc., in the compressed domain. A particular challenge is to understand how this paradigm can help exploiting truly multimedia features, which combine information from different associated streams such as audio and video, for joint audiovisual processing.

5. Highlights of the Year

5.1. Highlights of the Year

5.1.1. Awards

Antoine Deleforge (new PANAMA team member), Florence Forbes (MISTIS team) and Radu Horaud (PERCEPTION team) received the 2016 Hojjat Adeli Award for Outstanding Contributions in Neural Systems for their paper [70].

A. Deleforge, F. Forbes, and R. Horaud, "Acoustic space learning for sound-source separation and localization on binaural manifolds," *International journal of neural systems*, vol. 25, no. 01, 2015, https://hal.archives-ouvertes.fr/hal-00960796

The Award for Outstanding Contributions in Neural Systems established by World Scientific Publishing Co. in 2010, is awarded annually to the most innovative paper published in the previous volume/year of the International Journal of Neural Systems.

6. New Software and Platforms

6.1. Audio Activity Detector

KEYWORD: Audio activity estimation

- Authors: Frédéric Bimbot, Ewen Camberlein, Romain Lebarbenchon and Vincent Soupe
- Contact: Frédéric Bimbot

6.2. Audio Breath Rhythm Estimator

KEYWORD: Breath rhythm estimation

- Authors: Frédéric Bimbot, Ewen Camberlein and Romain Lebarbenchon
- Contact: Frédéric Bimbot

6.3. Audio GMM Classifier

- Authors: Frédéric Bimbot, Vincent Soupe, Jérémy Paret, Ewen Camberlein and Romain Lebarbenchon
- Contact: Frédéric Bimbot

6.4. CSCbox

Compressive Spectral Clustering Toolbox
KEYWORD: Clustering
SCIENTIFIC DESCRIPTION

The Compressive Spectral Clustering Toolbox is a Matlab toolbox implementing routines to reproduce experiments from the paper "Compressive Spectral Clustering", by N. Tremblay, G. Puy, P. Vandergheynst and R. Gribonval.
FUNCTIONAL DESCRIPTION

Matlab toolbox implementing routines to reproduce experiments from the paper "Compressive Spectral Clustering"

- Authors: Nicolas Tremblay, Gilles Puy, Pierre Vandergheynst and Rémi Gribonval
- Partner: EPFL - Ecole Polytechnique Fédérale de Lausanne
- Contact: Rémi Gribonval
- URL: http://www.irisa.fr/panama/software

6.5. FASST2

Flexible Audio Source Separation Toolbox
KEYWORDS: Audio - Source Separation
SCIENTIFIC DESCRIPTION

Only source separation software publicly availble allowing to use both spacial and spectral source properties with a generalised EM algorithm (expectation - maximisation). Fast specification of each use case by the choice of suitable constraints in constraint libraries.
FUNCTIONAL DESCRIPTION

Toolbox for the fast design of audio source separation adapted to any use case.

- Participants: Emmanuel Vincent and Yann Salaun
- Contact: Emmanuel Vincent
- URL: http://fasst.gforge.inria.fr

6.6. FAuST

KEYWORDS: Learning - Sparsity - Fast transform - Multilayer sparse factorisation
FUNCTIONAL DESCRIPTION

C++ toolbox, designed to decompose a given dense matrix into a product of sparse matrices in order to reduce its computational complexity (both for storage and manipulation).

- Authors: Luc Le Magoarou, Rémi Gribonval, Adrien Leman, Nicolas Bellot and Thomas Gautrais
- Contact: Rémi Gribonval
- URL: http://faust.gforge.inria.fr/

6.7. Multi-channel BSS Locate Basic

KEYWORDS: Audio - Localization - Signal processing - Multichannel signal

Scientific Description

Multi-Channel BSS Locate is a Matlab toolbox to estimate Direction Of Arrival (expressed both in azimuth and elevation) of multiple sources in a multi-channel audio signal recorded by an array of microphones. This toolbox implements the previous 8 angular spectrum methods presented in BSS Locate (GCC-PHAT, GCC-NONLIN, MUSIC and several SNR-based spectra).

- Authors: Emmanuel Vincent, Charles Blandin, Alexey Ozerov, Ewen Camberlein, Romain Lebarbenchon, Frédéric Bimbot and Nancy Bertin
- Contact: Emmanuel Vincent
- URL: http://bass-db.gforge.inria.fr/bss_locate/

6.8. SPADE

Sparse Audio Declipper
Keywords: Audio - Sparse regularization - Declipping
Scientific Description

Matlab routines to reproduce audio declipping experiments from the paper: - Srdan Kitic, Nancy Bertin, Remi Gribonval. Audio Declipping by Cosparse Hard Thresholding. iTwist - 2nd international - Traveling Workshop on Interactions between Sparse models and Technology, Aug 2014, Namur, Belgium. <hal-00922497v3> - Srdan Kitic, Nancy Bertin, Remi Gribonval. Sparsity and cosparsity for audio declipping: a flexible non-convex approach. LVA/ICA 2015 - The 12th International Conference on Latent Variable Analysis and Signal Separation, Aug 2015, Liberec, Czech Republic. pp.8. <hal-01159700v2>

- Participants: Srdan Kitic, Nancy Bertin and Rémi Gribonval
- Contact: Rémi Gribonval
- URL: http://xspaad.gforge.inria.fr/

6.9. SPOD Audio

Keywords: Audio source classification - Speaker verification - Breath rhythm estimation - Audio activity estimation

- Authors: Frédéric Bimbot, Vincent Soupe, Ewen Camberlein and Romain Lebarbenchon
- Contact: Frédéric Bimbot
- URL: http://www.kerlink.fr/en/

6.10. SPOD Model Generation

Keywords: Machine learning - Audio source classification - Statistical modeling - Speaker verification

- Authors: Frédéric Bimbot, Vincent Soupe, Jérémy Paret, Ewen Camberlein and Romain Lebarbenchon
- Contact: Frédéric Bimbot

6.11. SRP-PHAT

Keyword: Source localization

- Authors: Frédéric Bimbot, Nancy Bertin, Ewen Camberlein, Romain Lebarbenchon, Emmanuel Vincent, Charles Blandin and Alexey Ozerov
- Contact: Frédéric Bimbot

6.12. SketchMLBox

Keyword: Clustering

SCIENTIFIC DESCRIPTION

The SketchMLbox is a Matlab toolbox for fitting mixture models to large collections of training vectors using sketching techniques. The collection is first compressed into a vector called sketch, then a mixture model (e.g. a Gaussian Mixture Model) is estimated from this sketch using greedy algorithms typical of sparse recovery. The size of the sketch does not depend on the number of elements in the collection, but rather on the complexity of the problem at hand [2,3]. Its computation can be massively parallelized and distributed over several units. It can also be maintained in an online setting at low cost. Mixtures of Diracs ("K-means") and Gaussian Mixture Models with diagonal covariance are currently available, the toolbox is structured so that new mixture models can be easily implemented.

FUNCTIONAL DESCRIPTION

Matlab toolbox for fitting mixture models to large collections of feature vectors using sketching techniques.

- Authors: Nicolas Keriven, Rémi Gribonval and Nicolas Tremblay
- Partner: Université de Rennes 1
- Contact: Rémi Gribonval
- URL: http://sketchml.gforge.inria.fr

6.13. VoiceHome Corpus

KEYWORDS: Audio - Source Separation

FUNCTIONAL DESCRIPTION

This corpus includes reverberated, noisy speech signals spoken by native French talkers in a lounge and recorded by an 8-microphone device at various angles and distances and in various noise conditions. Room impulse responses and noise-only signals recorded in various real rooms and homes and baseline speaker localization and enhancement software are also provided.

- Contact: Nancy Bertin
- URL: http://voice-home.gforge.inria.fr/voiceHome_corpus.html

6.14. graphsamplingbox

- Authors: Nicolas Tremblay, Gilles Puy, Pierre Vandergheynst and Rémi Gribonval
- Partner: EPFL - Ecole Polytechnique Fédérale de Lausanne
- Contact: Rémi Gribonval
- URL: http://www.irisa.fr/panama/software

7. New Results

7.1. Recent results on Sparse Representations, Inverse Problems, and Dimension Reduction

Sparsity, low-rank, dimension-reduction, inverse problem, sparse recovery, scalability, compressive sensing

The team has had a substantial activity ranging from theoretical results to algorithmic design and software contributions in the fields of sparse representations, inverse problems, and dimension reduction, which is at the core of the ERC project PLEASE (Projections, Learning and Sparsity for Efficient Data Processing, see Section 9.2.1.1).

7.1.1. Theoretical results on Sparse Representations, Graph Signal Processing, and Dimension Reduction

Participants: Rémi Gribonval, Yann Traonmilin, Gilles Puy, Nicolas Tremblay, Pierre Vandergheynst.

Main collaboration: Mike Davies (University of Edinburgh), Pierre Borgnat (ENS Lyon), and members of the LTS2 lab of Pierre Vandergheynst at EPFL

Stable recovery of low-dimensional cones in Hilbert spaces: Many inverse problems in signal processing deal with the robust estimation of unknown data from underdetermined linear observations. Low dimensional models, when combined with appropriate regularizers, have been shown to be efficient at performing this task. Sparse models with the ℓ_1-norm or low rank models with the nuclear norm are examples of such successful combinations. Stable recovery guarantees in these settings have been established using a common tool adapted to each case: the notion of restricted isometry property (RIP). We established generic RIP-based guarantees for the stable recovery of cones (positively homogeneous model sets) with arbitrary regularizers. These guarantees were illustrated on selected examples. For block structured sparsity in the infinite dimensional setting, we used the guarantees for a family of regularizers which efficiency in terms of RIP constant can be controlled, leading to stronger and sharper guarantees than the state of the art. This has been published in a journal paper [21].

Recipes for stable linear embeddings from Hilbert spaces to $\mathbb{R}^m$: We considered the problem of constructing a linear map from a Hilbert space (possibly infinite dimensional) to $\mathbb{R}^m$ that satisfies a restricted isometry property (RIP) on an arbitrary signal model set. We obtained a generic framework that handles a large class of low-dimensional subsets but also *unstructured* and *structured* linear maps. We provided a simple recipe to prove that a random linear map satisfies a general RIP on the model set with high probability. We also described a generic technique to construct linear maps that satisfy the RIP. Finally, we detailed how to use our results in several examples, which allow us to recover and extend many known compressive sampling results. This has been presented at the conference EUSIPCO 2015 [90], and a journal paper is under revision [91].

Signal processing on graphs: from filtering to random sampling and robust PCA: Graph signal processing is an emerging field aiming at extending classical tools from signal processing (1D time series) and image processing (2D pixel grids, 3D voxel grids) to more loosely structured numerical data: collections of numerical values each associated to a vertex of a graph, where the graph encodes the underlying "topology" of proximities and distances. Since our pioneering contributions on this topic [4], the team regularly works on various aspects of graph signal processing, in collaboration with the LTS2 lab of Pierre Vandergheynst at EPFL. This year, we studied the problem of sampling k-bandlimited signals on graphs. We proposed two sampling strategies that consist in selecting a small subset of nodes at random. The first strategy is non-adaptive, i.e., independent of the graph structure, and its performance depends on a parameter called the graph coherence. On the contrary, the second strategy is adaptive but yields optimal results. Indeed, no more than $O(klog(k))$ measurements are sufficient to ensure an accurate and stable recovery of all k-bandlimited signals. This second strategy is based on a careful choice of the sampling distribution, which can be estimated quickly. Then, we proposed a computationally efficient decoder to reconstruct k-bandlimited signals from their samples. We proved that it yields accurate reconstructions and that it is also stable to noise. Finally, we conducted several experiments to test these techniques. A journal paper has been published [17] accompanied by a toolbox for reproducible research (see Section 6.14). Other contributions from this year on the topic of graph signal processing include new subgraph-based filterbanks for graph signals [22], and new accelerated and robustified techniques for PCA on graphs [19], [20] (see also below our contributions in terms of new algorithms to obtain approximate Fast Graph Fourier Transforms [32], [53]).

Accelerated spectral clustering: We leveraged the proposed random sampling technique to propose a faster spectral clustering algorithm. Indeed, classical spectral clustering is based on the computation of the first k eigenvectors of the similarity matrix' Laplacian, whose computation cost, even for sparse matrices, becomes prohibitive for large datasets. We showed that we can estimate the spectral clustering distance matrix without computing these eigenvectors: by graph filtering random signals. Also, we took advantage of the stochasticity of these random vectors to estimate the number of clusters k. We compared our method to classical spectral clustering on synthetic data, and showed that it reaches equal performance while being faster by a factor at least two for large datasets of real data. Two conference papers have been presented, at ICASSP 2016 [39] and ICML 2016 [40] and a toolbox for reproducible research has been released (see Section 6.4).

7.1.2. An Alternative Framework for Sparse Representations: Sparse "Analysis" Models

Participants: Rémi Gribonval, Nancy Bertin, Srdan Kitic, Clément Gaultier.

In the past decade there has been a great interest in a synthesis-based model for signals, based on sparse and redundant representations. Such a model assumes that the signal of interest can be composed as a linear combination of *few* columns from a given matrix (the dictionary). An alternative *analysis-based* model can be envisioned, where an analysis operator multiplies the signal, leading to a *cosparse* outcome.

Building on our pioneering work on the cosparse model [7] [73], [87] successful applications of this approach to sound source localization, audio declipping and brain imaging have been developed in 2015 and 2016. In addition, new applications to audio denoising were also introduced this year.

Versatile cosparse regularization: Digging the groove of previous years' results (comparison of the performance of several cosparse recovery algorithms in the context of sound source localization [77], demonstration of its efficiency in situations where usual methods fail ([79], see paragraph 7.4.2), applicability to the hard declipping problem [78], application to EEG brain imaging [56]), a journal paper embedding the latest algorithms and results in sound source localization and brain source localization in a unified fashion was published this year [5]. This framework was also exploited to extend results on audio inpainting (see Section 7.3.2).

New results include experimental confirmation of robustness and versatility of the proposed scheme, and of its computational merits (convergence speed increasing with the amount of data). In a work presented in a workshop [44], we also proposed a multiscale strategy that aims at exploiting computational advantages of both sparse and cosparse regularization approaches, thanks to the simple yet effective all-zero initialization which the synthesis-based optimization can benefit from, while retaining the computational properties of the analysis-based approach for huge scale optimization problems arising in physics-driven settings.

Parametric operator learning for cosparse calibration: In many inverse problems, a key challenge is to cope with unknown physical parameters of the problem such as the speed of sound or the boundary impedance. In the sound source localization problem, we previously showed that the unknown speed of sound can be learned jointly in the process of cosparse recovery, under mild conditions [58], [81]. This year, we extended the formulation to the case of unknown boundary impedance, and showing that a similar biconvex formulation and optimization could solve this new problem efficiently (conference paper published in ICASSP 2016 [29], see also Section 7.3.3).

7.1.3. Algorithmic and Theoretical results on Computational Representation Learning

Participants: Rémi Gribonval, Luc Le Magoarou, Nicolas Bellot, Adrien Leman, Cassio Fraga Dantas, Igal Rozenberg.

An important practical problem in sparse modeling is to choose the adequate dictionary to model a class of signals or images of interest. While diverse heuristic techniques have been proposed in the literature to learn a dictionary from a collection of training samples, classical dictionary learning is limited to small-scale problems. Inspired by usual fast transforms, we proposed a general dictionary structure that allows cheaper manipulation, and an algorithm to learn such dictionaries together with their fast implementation. The principle and its application to image denoising appeared at ICASSP 2015 [84] and an application to speedup linear inverse problems was published at EUSIPCO 2015 [83]. A Matlab library has been released (see Section 6.6) to reproduce the experiments from the comprehensive journal paper published this year [16], which additionally includes theoretical results on the improved sample complexity of learning such dictionaries. Pioneering identifiability results have been obtained in the Ph.D. thesis of Luc Le Magoarou on this topic [85].

We further explored the application of this technique to obtain fast approximations of Graph Fourier Transforms. A conference paper on this latter topic appeared in ICASSP 2016 [32], and a journal paper has been submitted [53] where we empirically show that $\mathcal{O}(n \log n)$ approximate implementations of Graph Fourier Transforms are possible for certain families of graphs. This opens the way to substantial accelerations for Fourier Transforms on large graphs.

A C++ software library has been developed (see Section 6.6) to release the resulting algorithms.

7.2. Activities on Waveform Design for Telecommunications

Peak to Average Power Ratio (PAPR), Orthogonal Frequency Division Multiplexing (OFDM), Generalized Waveforms for Multi Carrier (GWMC), Adaptive Wavelet Packet Modulation (AWPM)

7.2.1. Characterizing and designing multi-carrier waveform systems with optimum PAPR

Participant: Rémi Gribonval.

Main collaboration: Marwa Chafii, Jacques Palicot, Carlos Bader (Equipe SCEE, Supelec, Rennes)

In the context of the TEPN (Towards Energy Proportional Networks) Comin Labs project (see Section 9.1.1.2), in collaboration with the SCEE team at Supelec (thesis of Marwa Chafii [64], defended in October this year and co-supervised by R. Gribonval), we investigated a problem related to dictionary design: the characterization of waveforms with low Peak to Average Power Ratio (PAPR) for wireless communications. This is motivated by the importance of a low PAPR for energy-efficient transmission systems. A first stage of the work consisted in characterizing the statistical distribution of the PAPR for a general family of multi-carrier systems, leading to a journal paper [67] and several conference communications [65], [66]. Our characterization of waveforms with optimum PAPR [68] has been published in a journal this year [14]. The work this year has concentrated on designing new adaptive multi-carrier waveform systems able to cope with frequency-selective channels while minimizing PAPR. This has given rise to a patent [49] and a journal paper is in preparation.

7.3. Emerging activities on Compressive Learning and Nonlinear Inverse Problems

Compressive sensing, compressive learning, audio inpainting, phase estimation

7.3.1. Phase Estimation in Multichannel Mixtures

Participants: Antoine Deleforge, Yann Traonmilin.

The problem of estimating source signals given an observed multichannel mixture is fundamentally ill-posed when the mixing matrix is unknown or when the number of sources is larger that the number of microphones. Hence, prior information on the desired source signals must be incorporated in order to tackle it. An important line of research in audio source separation over the past decade consists in using a model of the source signals' magnitudes in the short-time Fourier domain [8]. Such models can be inferred through, *e.g.*, non-negative matrix factorization [89] or deep neural networks [88]. Magnitudes estimates are often interpreted as instantaneous variances of Gaussian-process source signals, and are combined with Wiener filtering for source separation. In [50], we introduced a shift of this paradigm by considering the *Phase Unmixing* problem: how can one recover the instantaneous phases of complex mixed source signals when their magnitudes and mixing matrix are known? This problem was showed to be NP-hard, and three approaches were proposed to tackle it: a heuristic method, an alternate minimization method, and a convex relaxation into a semi-definite program. The last two approaches were showed to outperform the oracle multichannel Wiener filter in under-determined informed source separation tasks. The latter yielded best results, including the potential for exact source separation in under-determined settings.

7.3.2. Audio Inpainting and Denoising

Participants: Rémi Gribonval, Nancy Bertin, Srdan Kitic.

Inpainting is a particular kind of inverse problems that has been extensively addressed in the recent years in the field of image processing. Building upon our previous pioneering contributions (definition of the audio inpainting problem as a general framework for many audio processing tasks, application to the audio declipping or desaturation problem, formulation as a sparse recovery problem [55]), we proposed over the last two years a series of algorithms leveraging the competitive cosparse approach, which offers a very appealing trade-off between reconstruction performance and computational time [78], [80], [81]. The work on cosparse audio declipping which was awarded the Conexant best paper award at the LVA/ICA 2015 conference [80], together with the associated toolbox for reproducible research (see Section 6.8) draw the attention of a world leading company in professional audio signal processing, with which some transfer has been negotiated. In 2016, real-time implementation of the A-SPADE algorithm was obtained and demonstrated at various events (HCERES evaluation, TechnofÃ©rence # 18 « Nouvelles expériences son et vidéo », ...).

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, denoising, time-frequency inpainting, joint source separation and declipping). In particular, we investigated the incorporation of the so-called "social" structure constraint [82] into problems regularized by a cosparse prior, including declipping and denoising. Publication of this work is currently under preparation.

7.3.3. Blind Calibration of Impedance and Geometry

Participants: Rémi Gribonval, Nancy Bertin, Srdan Kitic.

Main collaborations: Laurent Daudet, Thibault Nowakowski, Julien de Rosny (Institut Langevin)

Last year, we also investigated extended inverse problem scenarios where a "lack of calibration" may occur, i.e., when some physical parameters are needed for reconstruction but apriori unknown: speed of sound, impedance at the boundaries of the domain where the studied phenomenon propagates, or even the shape of these boundaries. In a first approach, based on our physics-driven cosparse regularization of the sound source localization problem [5] (see section 7.1.2), we managed to preserve the sound source localization performance when the speed of sound is unknown, or, equally, when the impedance is unknown, provided the shape is and under some smoothness assumptions. Unlike the previous case (gain calibration), the arising problems are not convex but biconvex, and can be solved with proper biconvex formulation of ADMM algorithm. In a second approach based on eigenmode decomposition (limited to a 2D membrane), we showed that impedance learning with known shape, or shape learning with known impedance can be expressed as two facets of the same problem, and solved by the same approach, from a small number of measurements. Two papers presenting these two sets of results appeared at ICASSP 2016 [29], [37].

7.3.4. Sketching for Large-Scale Mixture Estimation

Participants: Rémi Gribonval, Nicolas Keriven.

Main collaborations: Patrick Perez (Technicolor R&I France) Anthony Bourrier (formerly Technicolor R&I France, then GIPSA-Lab)

When fitting a probability model to voluminous data, memory and computational time can become prohibitive. We proposed during the Ph.D. thesis of Anthony Bourrier [60] a framework aimed at fitting a mixture of isotropic Gaussians to data vectors by computing a low-dimensional sketch of the data. The sketch represents empirical moments of the underlying probability distribution. Deriving a reconstruction algorithm by analogy with compressive sensing, we experimentally showed that it is possible to precisely estimate the mixture parameters provided that the sketch is large enough. The proposed algorithm provided good reconstruction and scaled to higher dimensions than previous probability mixture estimation algorithms, while consuming less memory in the case of voluminous datasets. It also provided a potentially privacy-preserving data analysis tool, since the sketch does not explicitly disclose information about individual datum it is based on [63], [61],

[62]. Last year, we consolidated our extensions to non-isotropic Gaussians, with new algorithms [76] and conducted large-scale experiments demonstrating its potential for speaker verification. A conference paper appeared at ICASSP 2016 [31] and a journal version has been submitted [52], accompanied by a toolbox for reproducible research (see Section 6.12).

This year the work concentrated on extending the approach beyond the case of Gaussian Mixture Estimation. First, we showed empirically that the algorithm can be adapted to sketch a training collection while still allowing to compute clusters. The approach, called “Compressive K-means”, is described in a paper accepted at ICASSP 2017 [27]. Then, we expressed a theoretical framework for sketched learning, encompassing statistical learning guarantees as well as dimension reduction guarantees. The framework already covers compressive K-means as well as compressive Principal Component Analysis (PCA), and a conference paper has been submitted. A comprehensive journal paper is under preparation, and future work will include expliciting the impact of the proposed framework on a wider set of concrete learning problems.

7.4. Source Separation and Localization

Source separation, sparse representations, probabilistic model, source localization

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 [8], which generalizes a number of existing techniques including our former study on spectral GMMs [57]. 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 [71], [75], [74], knowledge on the source positions and room characteristics [72], or a better initialization of parameters thanks to specific source localization techniques [59].

This accumulated progress lead, in 2015, to two main achievements: a new version of the Flexible Audio Source Separation Toolbox, fully reimplemented, was released [92] and we published an overview paper on recent and going research along the path of *guided* separation in a special issue of IEEE Signal Processing Magazine devoted to source separation and its applications [10]. This two achievements formed the basis of our work in 2016, exploring intensively the concrete use of these tools and principles in real-world scenarios, in particular within the voiceHome project (see Section 6.13).

7.4.1. Towards Real-world Separation and Remixing Applications

Participants: Nancy Bertin, Frédéric Bimbot, Ewen Camberlein, Romain Lebarbenchon.

In 2015, we began a new industrial collaboration, in the context of the VoiceHome project, aiming at another challenging real-world application: natural language dialog in home applications, such as control of domotic and multimedia devices. As a very noisy and reverberant environment, home is a particularly challenging target for source separation, used here as a pre-processing for speech recognition (and possibly with stronger interactions with voice activity detection or speaker identification tasks as well). In 2016, we publicly released a realistic corpus of room impulse responses and utterances recorded in real homes, and presented it during the Interspeech conference [28]. We also continued benchmarking and adapting existing localization and separation tools to the particular context of this application, worked on a better interface between source localization and source separations steps, and investigated new means to reduce the latency and computational burden of the currently available tools (low-resolution source separation preserving speech recognition improvement, automatic selection of the best microphones, joint localization and multichannel speech / non speech classification prior to any separation).

In november 2016, we started investigating a new application of source separation to sound respatialization from Higher Order Ambisonics (HOA) signals, in the context of free navigation in 3D audiovisual contents. This work is conducted in a collaboration with the IRT b<>Com, through the Ph.D. of Mohammed Hafsati (co-supervised by Nancy Bertin, RÃ©mi Gribonval).

7.4.2. Implicit Localization through Audio-based Control for Robotics

Participant: Nancy Bertin.

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*) [59]. 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, see Section 7.1.2).

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 investigate over the last three years 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 a 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 [69].

Last year, we established an analytical model linking the Interaural Time Difference (ITD) 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 and experiments with an actual robot [86]. This year, we consolidated these results and extended the range of applicative tasks [36] and obtained similar results (including theoretical and experimental) for the Interaural Level Difference (ILD), in combination with the absolute energy level [34]. Another set of experiments, presented during the IROS workshop [35] was successfully carried with a humanoid robot, notably without any measurement nor modeling of the robot's Head Relative Transfer Functions (HRTF). This work was mainly lead by Aly Magassouba, who defended his Ph.D. (co-supervised by Nancy Bertin and François Chaumette) in December 2016.

7.4.3. Emerging activities on Virtually-Supervised Sound Localization

Participants: Antoine Deleforge, Clément Gaultier, Saurabh Kataria.

Audio source localization consists in estimating the position of one or several sound sources given the signals received by a microphone array. It can be decomposed into two sub-tasks : (i) computing spatial auditory features from raw audio input and (ii) mapping these features to the desired spatial information.

Extracting spatial features from raw audio input: The most commonly used features in binaural (two microphones) sound source localization are frequency-dependent phase and level differences between the two microphones. To handle the presence of noise, several sources, or reverberation, most existing methods rely on some kind of aggregation of these features in the time-frequency plane, often in a heuristic way. In [25], we introduced the rectified binaural ratio as a new spatial feature. We showed that for Gaussian point-source signals corrupted by stationary Gaussian noise, this ratio follows a complex t-distribution with explicit parameters. This new formulation provides a principled, statistically sound and efficient method to aggregate these features in the presence of noise. Experiments notably showed the higher robustness of these features compared to traditional ones, in the task of localizing heavily corrupted speech signals.

Mapping features to spatial information: Existing methods to map auditory features to spatial properties divide into two categories. *Physics-driven* methods attempt to estimate an explicit mapping based on an approximate physical model of sound propagation in the considered system. *Data-driven* methods bypass the use of a physical model by learning the mapping from a training set, obtained by manually annotating features extracted from real data. We proposed a new paradigm that aims at making the best of physics-driven and data-driven approaches, referred to as *virtually-supervised acoustic space mapping* [26], [51]. The idea is to use a physics-based room-acoustic simulator to generate arbitrary large datasets of room-impulse responses corresponding to various acoustic environments, adapted to the physical audio system at hand. We demonstrated that mappings learned from these data could potentially be used to not only estimate the 3D position of a source but also some acoustical properties of the room [51]. We also showed that a virtually-learned mapping could robustly localize sound sources from real-world binaural input, which is the first result of this kind in audio source localization [26].

7.5. Music Content Processing and Information Retrieval

Music structure, music language modeling, System & Contrast model

Current work developed in our research group in the domain of music content processing and information retrieval explore various information-theoretic frameworks for music structure analysis and description [24], in particular the System & Contrast model [1].

7.5.1. Tensor-based Representation of Sectional Units in Music

Participants: Corentin Guichaoua, Frédéric Bimbot.

Following Kolmogorov'ÂÂs complexity paradigm, modeling the structure of a musical segment can be addressed by searching for the compression program that describes as economically as possible the musical content of that segment, within a given family of compression schemes.

In this general framework, packing the musical data in a tensor-derived representation enables to decompose the structure into two components : (i) the shape of the tensor which characterizes the way in which the musical elements are arranged in an n-dimensional space and (ii) the values within the tensor which reflect the content of the musical segment and minimize the complexity of the relations between its elements.

This approach is currently developed and tested for the grouping of chord sequences into sectional units for pop music songs, with very encouraging segmentation results on pop songs.

7.5.2. Minimal Transport Graphs for the Modeling of Chord Progressions

Participants: Corentin Louboutin, Frédéric Bimbot.

In this work, we model relations between chords by minimal transport and we investigate different types of dependencies within chord sequences [33]. For this purpose we use the ÂÂSystem & Contrast (S&C) model [1], designed for the description of music sectional units, to infer non-sequential structures called chord progression graphs (CPG).

Minimal transport is defined as the shortest displacement of notes, in semitones, between a pair of chords. The paper [33] present three algorithms to find CPGs for chords sequences: one is sequential, and two others are based on the S& C model. The three methods are compared using the perplexity as an efficiency measure.

The experiments on a corpus of 45 segments taken from songs of multiple genres indicate that optimization processes based on the S&C model outperform the sequential model with a decrease in perplexity over 1.0.

7.5.3. Regularity Constraints for the Fusion of Music Structure Segmentation System

Participant: Frédéric Bimbot.

Main collaborations Gabriel Sargent (EPI LinkMedia, Rennes, France)

Music structure estimation has recently emerged as a central topic within the field of Music Information Retrieval. Indeed, as music is a highly structured information stream, knowledge of how a music piece is organized represents a key challenge to enhance the management and exploitation of large music collections.

Former work carried out in our group [9] has illustrated the benefits that can be expected from a regularity constraint on the structural segmentation of popular music pieces : a constraint which favors structural segments of comparable size provides a better conditioning of the boundary estimation process.

As a further investigation, we have explored the benefits of the regularity constraint as an efficient way for combining the outputs of a selection of systems presented at MIREX between 2010 and 2015. These experiments have yielded a level of performance which is competitive to that of the state-of-the-art on the ”MIREX10" dataset (100 J-Pop songs from the RWC database) [18].

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Contracts with Industry

8.1.1. Licensing agreement contract with Cedar Audio Limited

Participants: Nancy Bertin, Srdan Kitic, Rémi Gribonval.

This contract aimed at licensing an audio desaturation (declipping) software developed in the team.

8.2. Bilateral Grants with Industry

8.2.1. CIFRE contract with Technicolor R&I France on Very large scale visual comparison

Participants: Rémi Gribonval, Himalaya Jain.

Duration: 3 years (2015-2018)

Research axis: 3.1.2

Partners: Technicolor R&I France, Inria-Rennes

Funding: Technicolor R&I France, ANRT

The grand goal of this thesis is to design, analyze and test new tools to allow large-scale comparison of high-dimensional visual signatures. Leveraging state of the art visual descriptors, the objective is to obtain new compact codes for visual representations, exploiting sparsity and learning, so that they can be stored and compared in an efficient, yet meaningful, way.

9. Partnerships and Cooperations

9.1. National Initiatives

9.1.1. Labex Comin Labs projects

CominLabs is a Laboratoire d'Excellence funded by the PIA (Programme Investissements d'Avenir) in the broad area of telecommunications.

9.1.1.1. HEMISFER

Participant: Rémi Gribonval.

Acronym: HYBRID (Hybrid Eeg-MrI and Simultaneous neuro-feedback for brain Rehabilitation)
http://www.hemisfer.cominlabs.ueb.eu/
Research axis: 3.1
CominLabs partners : EPI VISAGES; EPI HYBRID; EPI PANAMA
External partners : EA 4712 team from University of Rennes I; EPI ATHENA, Sophia-Antipolis;
Coordinator: Christian Barillot, EPI VISAGES
Description: The goal of HEMISFER is to make full use of neurofeedback paradigm in the context of rehabilitation and psychiatric disorders. The major breakthrough will come from the use of a coupling model associating functional and metabolic information from Magnetic Resonance Imaging (fMRI) to Electro-encephalography (EEG) to "enhance" the neurofeedback protocol. We propose to combine advanced instrumental devices (Hybrid EEG and MRI platforms), with new man-machine interface paradigms (Brain computer interface and serious gaming) and new computational models (source separation, sparse representations and machine learning) to provide novel therapeutic and neuro-rehabilitation paradigms in some of the major neurological and psychiatric disorders of the developmental and the aging brain (stroke, attention-deficit disorder, language disorders, treatment-resistant mood disorders, Ã¢ÂÂ¦).
Contribution of PANAMA: PANAMA, in close cooperation with the VISAGES team, contributes to a coupling model between EEG and fMRI considered as a joint inverse problem addressed with sparse regularization. By combining both modalities, one expects to achieve a good reconstruction both in time and space. This new imaging technique will then be used for improving neurofeedback paradigms in the context of rehabilitation and psychiatric disorders, which is the final purpose of the HEMISFER project.

9.1.1.2. TEPN

Participant: Rémi Gribonval.

Acronym: TEPN (Toward Energy Proportional Networks)
http://www.tepn.cominlabs.ueb.eu/
Research axis: 3.1
CominLabs partners : IRISA OCIF - Telecom Bretagne; IETR SCN; IETR SCEE; EPI PANAMA
Coordinator: Nicolas Montavont, IRISA OCIF - Telecom Bretagne
Description: As in almost all areas of engineering in the past several decades, the design of computer and network systems has been aimed at delivering maximal performance without regarding to the energy efficiency or the percentage of resource utilization. The only places where this tendency was questioned were battery-operated devices (such as laptops and smartphones) for which the users accept limited (but reasonable) performance in exchange for longer use periods. Even though the end users make such decisions on a daily basis by checking their own devices, they have no way of minimizing their energy footprint (or conversely, optimize the network resource usage) in the supporting infrastructure. Thus, the current way of dimensioning and operating the infrastructure supporting the user services, such as cellular networks and data centers, is to dimension for peak usage. The problem with this approach is that usage is rarely at its peak. The overprovisioned systems are also aimed at delivering maximal performance, with energy efficiency being considered as something desired, but non-essential. This project aims at making the network energy consumption proportional to the actual charge of this network (in terms of number of served users, or requested bandwidth). An energy proportional network can be designed by taking intelligent decisions (based on various constraints and metrics) into the network such as switching on and off network components in order to adapt the energy consumption to the user needs. This concept can be summarized under the general term of Green Cognitive Network Approach.
Contribution of PANAMA: PANAMA, in close cooperation with the SCEE team at IETR (thesis of Marwa Chafii), focuses on the design of new waveforms for multi carrier systems with reduced Peak to Average Power Ratio (PAPR).

9.1.2. ANR INVATE project with IRT b<>com France

Participants: Rémi Gribonval, Nancy Bertin, Mohamed Hafsati.

Thesis on 3D audio scene decomposition for interactive navigation

Duration: 3 years (2016-2019)

Research axis: 3.2.2

Partners: IRT b<>com, Inria-Rennes, IRISA

Funding: ANR INVATE project (PIA)

The objective of this thesis is to develop tools to analyze audio scenes in order to identify, locate, and extract the sources present in the scene to re-spatialize them according to the user head orientation and the movement of the user in the targeted virtual scene.

9.1.3. OSEO-FUI: voiceHome

Participants: Nancy Bertin, Frédéric Bimbot, Romain Lebarbenchon, Ewen Camberlein.

Duration: 3 years (2015-2017)

Research axis: 3.2

Partners: onMobile, Delta Dore, eSoftThings, Orange, Technicolor, LOUSTIC, Inria Nancy

Coordinator: onMobile

Description: The goal of the project is to design and implement a multi-channel voice interface for smart home and multimedia (set-top-box) appliances.

Contributions of PANAMA are focused on (i) audio activity monitoring and wake-up word detection and (ii) audio source localization and separation. In both cases, the issue of energy frugality is central and strongly constrains the available resources. We expect from this cooperation to make progress towards operational low-resource audio source separation schemes and we intend to investigate compressive sensing for the characterization of audio and voice activity.

9.2. European Initiatives

9.2.1. FP7 & H2020 Projects

9.2.1.1. ERC-StG: PLEASE (Projections, Learning, and Sparsity for Efficient Data Processing)

Participants: Rémi Gribonval, Srdan Kitic, Luc Le Magoarou, Nancy Bertin, Nicolas Keriven, Yann Traonmilin, Gilles Puy, Adrien Leman, Nicolas Bellot.

Duration: January 2012 - December 2016

Research axis: 3.1

Principal investigator: Rémi Gribonval

Program: ERC Starting Grant

Project acronym: PLEASE

Project title: Projections, Learning and Sparsity for Efficient data processing

Abstract: The Please ERC is focused on the extension of the sparse representation paradigm towards that of sparse modeling, with the challenge of establishing, strengthening and clarifying connections between sparse representations and machine learning

Web site: https://team.inria.fr/panama/projects/please/

9.3. International Initiatives

9.3.1. Inria International Partners

9.3.1.1. Informal International Partners

PANAMA has strong recurrent collaborations with the LTS2 lab at EPFL, the Center for Digital Music at Queen Mary University of London, the Institute for Digital Communications at the University of Edimburgh, and the Institute for Mathematics of the Postdam University.

9.4. International Research Visitors

9.4.1. Visits of International Scientists

- Pierre Vandergheynst, in June-July, Professor of Signal and Image Processing, EPFL (Chaire Internationale Inria)
- Gilles Blanchard, in September, Professor, University of Potsdam
- Laurent Jacques, in October, Professor, Catholic University of Louvain
- Mike Davies, in November, Professor, University of Edimburgh

10. Dissemination

10.1. Promoting Scientific Activities

Rémi Gribonval is a member of the IEEE Technical Committee on Signal Processing Theory and Methods (2012–2017), and a member of the Awards sub-committee.

Rémi Gribonval is a member of the program committee of the GRETSI.

Rémi Gribonval is a member of the Steering Committee of the SPARS international workshop (chairman until 2013).

Frédéric Bimbot is the Head of the "Digital Signals and Images, Robotics" department in IRISA (UMR 6074).

Frédéric Bimbot is a member of the International Advisory Council of ISCA (International Speech Communication Association).

Rémi Gribonval and Frédéric Bimbot are the scientific coordinators of the Science and Music Day (Journée Science et Musique) organized by IRISA.

Antoine Deleforge organized and co-chaired with Prof. Sharon Gannot (Bar-Ilan University) a special session on “Learning-Based Sound Source Localization and Spatial Information Retrieval” at ICASSP 2016, Shanghai, China.

N. Bertin, A. Deleforge, and R. Gribonval co-organized a special session entitled “From Source Position to Room Properties : Learning Methods for Audio Scene Geometry Estimation”, at the 2017 LVA/ICA Conference, Grenoble, February 2017

R. Gribonval was the organizer of a special session on “Multiscale Factorizations and Learning”, at the 2016 IEEE Information Theory Workshop, Cambridge, UK, September 2016

R. Gribonval was guest editor of a special issue of IEEE Journal on Selected Topics in Signal Processing, on Structured Matrices in Signal and Data Processing.

F. Bimbot was an invited speaker at the Dagstuhl Seminar 16092 on Computational Music Structure Analysis (February 2016, Dagstuhl, Germany)

R. Gribonval was an invited speaker at the London Workshop on Sparse Signal Processing (Imperial College, September 2016), the DALI (Data Learning and Inference) Workshop on Learning Theory (Sestri Levante, March 2016), the Workshop on Computational and Statistical Trade-offs in Learning (IHES, Bures-sur-Yvette, March 2016), and the Workshop on Low Complexity Models in Signal Processing (Hausdorff Institute, Bonn, February 2016).

R. Gribonval is a member of the jury of the GDR ISIS / GRETSI / Club EAA thesis prize in signal and image processing since 2014.

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

Bachelor : N. Bertin, "Discovery of selected topics in audio signal processing research", 6 hours, L3, École Supérieure de Réalisation Audiovisuelle (ESRA), France.

Master : N. Bertin, "Audio rendering, coding and source separation", 6 hours, M2, Université Rennes 1, France.

Master : N. Bertin, "Audio indexing and classification", 6 hours, M2, Université Rennes 1, France.

Master : N. Bertin, "Fundamentals of Signal Processing", 24 hours, M1, Ecole Normale Supérieure de Bretagne, Rennes, France.

Master : N. Bertin, "Sparse representations and compressive sensing", 15 hours, M2, Ecole Nationale de la Statistique et de l'Analyse de l'Information, Rennes, France.

Master : N. Bertin, "Sparsity in Signal and Image Processing", 10 hours, M1, Ecole Normale Supérieure de Bretagne, Rennes, France.

Master : N. Bertin, "Sparsity in Signal and Image Processing", 12 hours, M2, Institut National des Sciences Appliquées (INSA) de Rennes, France.

Master : R. Gribonval, "Signal and image representations", 8 hours, M2, Université Rennes 1, France.

Master: R. Gribonval, coordination of the ARD module "Acquisition et Représensation de Données", 20hours, M2, Université Rennes 1, France.

Master : R. Gribonval, "Sparsity in Signal and Image Processing", 6 hours, M2, Institut National des Sciences Appliquées (INSA) de Rennes, France.

Master : R. Gribonval, coordination of the module "Sparsity in Signal and Image Processing", 48 hours, M2, Institut National des Sciences Appliquées (INSA) de Rennes, France.

Bachelor : A. Deleforge, "Discovery of selected topics in audio signal processing research", 6 hours, L3, École Supérieure de Réalisation Audiovisuelle (ESRA), France.

Master : A. Deleforge, "Sparsity in Signal and Image Processing", 4 hours, M2, Institut National des Sciences Appliquées (INSA) de Rennes, France.

10.3. Popularization

10.3.1. Journée Science et Musique

Participants: Antoine Deleforge, Rémi Gribonval, Frédéric Bimbot, Romain Lebarbenchon, Adrien Leman, Nicolas Bellot, Clément Gaultier, Nancy Bertin, Nicolas Keriven, Ewen Camberlein, Luc Le Magoarou, Armelle Mozziconacci, Stéphanie Lemaile, Corentin Louboutin, Nicolas Tremblay, Corentin Guichaoua, Srdan Kitic.

with contributions and support from: Valérie Gouranton, Laurent Perraudeau, Nathalie Denis, Evelyne Orain, Agnès Cottais, and many more.

PANAMA coordinated the organization of a public event called “JournÃ©e Science et Musique” (“Music and Science Day”). This yearly event organized by the METISS/ PANAMA Team since 2011 aims at sharing with the wide audience the latest innovations and research projects in music. The motivation for hosting this event is to explain and promote the technology behind audio-processing that people face in their daily lives. The event is free to everyone and people have the possibility to attend talks by selected speakers or meet numerous experts that demonstrate current projects in which people can interactively participate. Edition 2016 hosted more than 500 visitors and was the official opening event of the “Festival des Sciences” week in Rennes.

11. Bibliography

Major publications by the team in recent years

[1] F. Bimbot, E. Deruty, G. Sargent, E. Vincent.*System & Contrast : A Polymorphous Model of the Inner Organization of Structural Segments within Music Pieces*, in "Music Perception", 2016, 41, To appear - http://mp.ucpress.edu/, https://hal.inria.fr/hal-01188244.

[2] N. Duong, E. Vincent, R. Gribonval.*Under-determined reverberant audio source separation using a full-rank spatial covariance model*, in "IEEE Transactions on Audio, Speech and Language Processing", July 2010, vol. 18, n^{o} 7, p. 1830–1840 [*DOI :* 10.1109/TASL.2010.2050716], https://hal.inria.fr/inria-00541865.

[3] R. Gribonval, K. Schnass.*Dictionary Identification - Sparse Matrix-Factorisation via ℓ_1-Minimisation*, in "IEEE Transactions on Information Theory", July 2010, vol. 56, n^{o} 7, p. 3523–3539 [*DOI :* 10.1109/TIT.2010.2048466], https://hal.archives-ouvertes.fr/hal-00541297.

[4] D. K. Hammond, P. Vandergheynst, R. Gribonval.*Wavelets on graphs via spectral graph theory*, in "Applied and Computational Harmonic Analysis", March 2011, vol. 30, n^{o} 2, p. 129–150 [*DOI :* 10.1016/j.acha.2010.04.005], https://hal.inria.fr/inria-00541855.

[5] S. Kitić, L. Albera, N. Bertin, R. Gribonval.*Physics-driven inverse problems made tractable with cosparse regularization*, in "IEEE Transactions on Signal Processing", January 2016, vol. 64, n^{o} 2, p. 335-348 [*DOI :* 10.1109/TSP.2015.2480045], https://hal.inria.fr/hal-01133087.

[6] S. Kitić.*Cosparse regularization of physics-driven inverse problems*, IRISA, Inria Rennes, November 2015, https://hal.archives-ouvertes.fr/tel-01237323.

[7] S. Nam, M. E. Davies, M. Elad, R. Gribonval.*The Cosparse Analysis Model and Algorithms*, in "Applied and Computational Harmonic Analysis", 2013, vol. 34, n^{o} 1, p. 30–56, Preprint available on arXiv since 24 Jun 2011 [*DOI :* 10.1016/j.acha.2012.03.006], http://hal.inria.fr/inria-00602205.

[8] A. Ozerov, E. Vincent, F. Bimbot.*A General Flexible Framework for the Handling of Prior Information in Audio Source Separation*, in "IEEE Transactions on Audio, Speech and Language Processing", May 2012, vol. 20, n^{o} 4, p. 1118 - 1133, 16, http://hal.inria.fr/hal-00626962.

[9] G. Sargent.*Music structure estimation using multi-criteria analysis and regularity constraints*, Université Rennes 1, February 2013, https://tel.archives-ouvertes.fr/tel-00853737.

[10] E. Vincent, N. Bertin, R. Gribonval, F. Bimbot.*From blind to guided audio source separation*, in "IEEE Signal Processing Magazine", December 2013, http://hal.inria.fr/hal-00922378.

Publications of the year

Doctoral Dissertations and Habilitation Theses

[11] M. CHAFII. *Study of a new multicarrier waveform with low PAPR*, CentraleSupélec, October 2016, https://hal.archives-ouvertes.fr/tel-01399509.

Articles in International Peer-Reviewed Journal

[12] H. BECKER, L. ALBERA, P. COMON, R. GRIBONVAL, F. WENDLING, I. MERLET. *Localization of Distributed EEG Sources in the Context of Epilepsy: A Simulation Study*, in "IRBM", 2016 [*DOI :* 10.1016/J.IRBM.2016.04.001], https://hal.archives-ouvertes.fr/hal-01359237.

[13] F. BIMBOT, E. DERUTY, G. SARGENT, E. VINCENT. *System & Contrast : A Polymorphous Model of the Inner Organization of Structural Segments within Music Pieces*, in "Music Perception", 2016, 41, https://hal.inria.fr/hal-01188244.

[14] M. CHAFII, J. PALICOT, R. GRIBONVAL, F. BADER. *A Necessary Condition for Waveforms with Better PAPR than OFDM*, in "IEEE Transactions on Communications", 2016 [*DOI :* 10.1109/TCOMM.2016.2584068], https://hal.inria.fr/hal-01128714.

[15] S. KITIĆ, L. ALBERA, N. BERTIN, R. GRIBONVAL. *Physics-driven inverse problems made tractable with cosparse regularization*, in "IEEE Transactions on Signal Processing", January 2016, vol. 64, n^o 2, p. 335-348 [*DOI :* 10.1109/TSP.2015.2480045], https://hal.inria.fr/hal-01133087.

[16] L. LE MAGOAROU, R. GRIBONVAL. *Flexible Multi-layer Sparse Approximations of Matrices and Applications*, in "IEEE Journal of Selected Topics in Signal Processing", June 2016 [*DOI :* 10.1109/JSTSP.2016.2543461], https://hal.inria.fr/hal-01167948.

[17] G. PUY, N. TREMBLAY, R. GRIBONVAL, P. VANDERGHEYNST. *Random sampling of bandlimited signals on graphs*, in "Applied and Computational Harmonic Analysis", June 2016 [*DOI :* 10.1016/J.ACHA.2016.05.005], https://hal.inria.fr/hal-01229578.

[18] G. SARGENT, F. BIMBOT, E. VINCENT. *Estimating the structural segmentation of popular music pieces under regularity constraints*, in "IEEE/ACM Transactions on Audio, Speech, and Language Processing", 2017, https://hal.inria.fr/hal-01403210.

[19] N. SHAHID, N. PERRAUDIN, V. KALOFOLIAS, G. PUY, P. VANDERGHEYNST. *Fast Robust PCA on Graphs*, in "IEEE Journal of Selected Topics in Signal Processing", April 2016, vol. 10, n^o 4, p. 740 - 756 [*DOI :* 10.1109/JSTSP.2016.2555239], https://hal.inria.fr/hal-01277624.

[20] N. SHAHID, N. PERRAUDIN, G. PUY, P. VANDERGHEYNST. *Compressive PCA for Low-Rank Matrices on Graphs*, in "IEEE transactions on Signal and Information Processing over Networks", 2016, 17, Titled changed from initial preprint "Compressive PCA on graphs" [*DOI :* 10.1109/TSIPN.2016.2631890], https://hal.inria.fr/hal-01277625.

[21] Y. TRAONMILIN, R. GRIBONVAL. *Stable recovery of low-dimensional cones in Hilbert spaces: One RIP to rule them all*, in "Applied and Computational Harmonic Analysis", September 2016, https://hal.inria.fr/hal-01207987.

[22] N. TREMBLAY, P. BORGNAT.*Subgraph-based filterbanks for graph signals*, in "IEEE Transactions on Signal Processing", March 2016, https://hal.archives-ouvertes.fr/hal-01243889.

[23] M. B. WAKIN, R. GRIBONVAL, V. KOIVUNEN, J. ROMBERG, J. WRIGHT.*Introduction to the Issue on Structured Matrices in Signal and Data Processing*, in "IEEE Journal of Selected Topics in Signal Processing", May 2016, vol. 10, n^o 4, p. 605–607 [*DOI :* 10.1109/JSTSP.2016.2553398], https://hal.inria.fr/hal-01421797.

Invited Conferences

[24] F. BIMBOT.*Towards an Information-Theoretic Framework for Music Structure*, in "Dagstuhl Seminar on Computational Music Structure Analysis", Dagstuhk, Germany, February 2016, p. 167-168 [*DOI :* 10.4230/DAGREP.6.2.147], https://hal.archives-ouvertes.fr/hal-01421013.

[25] A. DELEFORGE, F. FORBES.*Rectified binaural ratio: A complex T-distributed feature for robust sound localization*, in "European Signal Processing Conference", Budapest, Hungary, August 2016, p. 1257-1261, https://hal.inria.fr/hal-01372337.

[26] C. GAULTIER, S. KATARIA, A. DELEFORGE.*VAST : The Virtual Acoustic Space Traveler Dataset*, in "International Conference on Latent Variable Analysis and Signal Separation (LVA/ICA)", Grenoble, France, February 2017, https://hal.archives-ouvertes.fr/hal-01416508.

[27] N. KERIVEN, N. TREMBLAY, Y. TRAONMILIN, R. GRIBONVAL.*Compressive K-means*, in "International Conference on Acoustics, Speech and Signal Processing (ICASSP)", New Orleans, United States, March 2017, https://hal.inria.fr/hal-01386077.

International Conferences with Proceedings

[28] N. BERTIN, E. CAMBERLEIN, E. VINCENT, R. LEBARBENCHON, S. PEILLON, É. LAMANDÉ, S. SIVASANKARAN, F. BIMBOT, I. ILLINA, A. TOM, S. FLEURY, E. JAMET.*A French corpus for distant-microphone speech processing in real homes*, in "Interspeech 2016", San Francisco, United States, September 2016, https://hal.inria.fr/hal-01343060.

[29] N. BERTIN, S. KITIĆ, R. GRIBONVAL.*Joint estimation of sound source location and boundary impedance with physics-driven cosparse regularization*, in "ICASSP - 41st IEEE International Conference on Acoustics, Speech and Signal Processing", Shanghai, China, March 2016, https://hal.archives-ouvertes.fr/hal-01247227.

[30] H. JAIN, P. PÉREZ, R. GRIBONVAL, J. ZEPEDA, H. JÉGOU.*Approximate search with quantized sparse representations*, in "14th European Conference on Computer Vision (ECCV)", Amsterdam, Netherlands, October 2016, https://hal.archives-ouvertes.fr/hal-01361953.

[31] N. KERIVEN, A. BOURRIER, R. GRIBONVAL, P. PÉREZ.*Sketching for Large-Scale Learning of Mixture Models*, in "2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2016)", Shanghai, China, March 2016, https://hal.inria.fr/hal-01208027.

[32] L. LE MAGOAROU, R. GRIBONVAL.*Are There Approximate Fast Fourier Transforms On Graphs?*, in "International Conference on Acoustics, Speech and Signal Processing (ICASSP)", Shanghai, China, March 2016, https://hal.archives-ouvertes.fr/hal-01254108.

[33] C. LOUBOUTIN, F. BIMBOT.*Description of Chord Progressions by Minimal Transport Graphs Using the System & Contrast Model*, in "ICMC 2016 - 42nd International Computer Music Conference", Utrecht, Netherlands, September 2016, https://hal.archives-ouvertes.fr/hal-01421023.

[34] A. MAGASSOUBA, N. BERTIN, F. CHAUMETTE.*Audio-based robot controlfrom interchannel level difference and absolute sound energy*, in "IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, IROS'16", Daejeon, South Korea, October 2016, p. 1992-1999, https://hal.inria.fr/hal-01355394.

[35] A. MAGASSOUBA, N. BERTIN, F. CHAUMETTE.*Binaural auditory interaction without HRTF for humanoid robots: A sensor-based control approach*, in "Workshop on Multimodal Sensor-based Control for HRI and soft manipulation, IROS'2016", Daejeon, South Korea, October 2016, https://hal.inria.fr/hal-01408422.

[36] A. MAGASSOUBA, N. BERTIN, F. CHAUMETTE.*First applications of sound-based control on a mobile robot equipped with two microphones*, in "IEEE Int. Conf. on Robotics and Automation, ICRA'16", Stockholm, Sweden, May 2016, https://hal.inria.fr/hal-01277589.

[37] T. NOWAKOWSKI, N. BERTIN, R. GRIBONVAL, J. DE ROSNY, L. DAUDET.*Membrane Shape And Boundary Conditions Estimation Using Eigenmode Decomposition*, in "ICASSP 2016 – 41st IEEE International Conference on Acoustics, Speech and Signal Processing", Shanghai, China, March 2016, https://hal.inria.fr/hal-01254882.

[38] A. SCHMIDT, A. DELEFORGE, W. KELLERMANN.*Ego-Noise Reduction Using a Motor Data-Guided Multichannel Dictionary*, in "International Conference on Intelligent Robots and Systems (IROS), 2016", Daejon, South Korea, IEEE/RSJ, October 2016, p. 1281-1286, https://hal.inria.fr/hal-01415723.

[39] N. TREMBLAY, G. PUY, P. BORGNAT, R. GRIBONVAL, P. VANDERGHEYNST.*Accelerated spectral clustering using graph filtering of random signals*, in "41st IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2016)", Shanghai, China, March 2016, https://hal.archives-ouvertes.fr/hal-01243682.

[40] N. TREMBLAY, G. PUY, R. GRIBONVAL, P. VANDERGHEYNST.*Compressive Spectral Clustering*, in "33rd International Conference on Machine Learning", New York, United States, June 2016, https://hal.archives-ouvertes.fr/hal-01320214.

National Conferences with Proceeding

[41] M. CHAFII, J. PALICOT, R. GRIBONVAL. *La modulation en ondelettes: une modulation alternative à faible consommation d'énergie*, in "URSI Énergie et radiosciences", Rennes, France, Journées Scientifiques URSI "Énergie et radiosciences", March 2016, n^{o} 143, https://hal.archives-ouvertes.fr/hal-01401604.

[42] N. MONTAVONT, D. SHEHADEH, J. PALICOT, X. LAGRANGE, A. BLANC, R. GRIBONVAL, P. MARY, J.-Y. BAUDAIS, J.-F. HÉLARD, M. CRUSSIÈRE, Y. LOUËT, C. MOY.*Toward Energy Proportional Networks*, in "Journées scientifiques, URSI 2016", Rennes, France, March 2016, https://hal.archives-ouvertes.fr/hal-01297894.

Conferences without Proceedings

[43] M. CHAFII, M. LAMARANA DIALLO, J. PALICOT, F. BADER, R. GRIBONVAL.*Adaptive Tone Reservation for better BER Performance in a Frequency Selective Fading Channel*, in "IEEE VTC2016-Spring", Nanjing, China, May 2016, https://hal-centralesupelec.archives-ouvertes.fr/hal-01290918.

[44] S. KITIC, N. BERTIN, R. GRIBONVAL.*The best of both worlds: synthesis-based acceleration for physics-driven cosparse regularization*, in "iTwist 2016 - International Traveling Workshop on Interactions Between Sparse Models and Technology", Aalborg, Denmark, August 2016, https://hal.archives-ouvertes.fr/hal-01329051.

[45] E. PERTHAME, F. FORBES, B. OLIVIER, A. DELEFORGE.*Non linear robust regression in high dimension*, in "The XXVIIIth International Biometric Conference", Victoria, Canada, July 2016, https://hal.archives-ouvertes.fr/hal-01423622.

[46] E. PERTHAME, F. FORBES, B. OLIVIER, A. DELEFORGE.*Regression non lineaire robuste en grande dimension*, in "48èmes Journées de Statistique organisées par la Société Française de Statistique", Montpellier, France, May 2016, https://hal.archives-ouvertes.fr/hal-01423630.

Research Reports

[47] C. LOUBOUTIN, F. BIMBOT.*Tensorial Description of Chord Progressions: Complementary Scientific Material*, IRISA, équipe PANAMA, May 2016, https://hal.archives-ouvertes.fr/hal-01314493.

[48] G. SARGENT, F. BIMBOT, E. VINCENT.*Supplementary material to the article: Estimating the structural segmentation of popular music pieces under regularity constraints*, IRISA-Inria, Campus de Beaulieu, 35042 Rennes cedex ; Inria Nancy, équipe Multispeech, September 2016, https://hal.inria.fr/hal-01368683.

Patents and standards

[49] M. CHAFII, J. PALICOT, R. GRIBONVAL.*Dispositif de communication à modulation temps-fréquence adaptative*, July 2016, n° Numéro de demande : 1656806 ; Numéro de soumission : 1000356937, https://hal.inria.fr/hal-01375661.

Other Publications

[50] A. DELEFORGE, Y. TRAONMILIN.*Phase Unmixing : Multichannel Source Separation with Magnitude Constraints*, September 2016, working paper or preprint, https://hal.inria.fr/hal-01372418.

[51] S. KATARIA, C. GAULTIER, A. DELEFORGE.*Hearing in a shoe-box : binaural source position and wall absorption estimation using virtually supervised learning* , September 2016, working paper or preprint, https://hal.inria.fr/hal-01372435.

[52] N. KERIVEN, A. BOURRIER, R. GRIBONVAL, P. PÉREZ.*Sketching for Large-Scale Learning of Mixture Models*, June 2016, working paper or preprint, https://hal.inria.fr/hal-01329195.

[53] L. LE MAGOAROU, R. GRIBONVAL, N. TREMBLAY.*Approximate Fast Fourier Transforms on graphs via multi-layer sparse approximations*, December 2016, working paper or preprint, https://hal.inria.fr/hal-01416110.

[54] E. PERTHAME, F. FORBES, A. DELEFORGE.*Inverse regression approach to robust non-linear high-to-low dimensional mapping*, July 2016, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01347455.

References in notes

[55] A. ADLER, V. EMIYA, M. G. JAFARI, M. ELAD, R. GRIBONVAL, M. D. PLUMBLEY.*Audio Inpainting*, in "IEEE Transactions on Audio, Speech and Language Processing", March 2012, vol. 20, n^{o} 3, p. 922 - 932 [*DOI :* 10.1109/TASL.2011.2168211], http://hal.inria.fr/inria-00577079.

[56] L. ALBERA, S. KITIĆ, N. BERTIN, G. PUY, R. GRIBONVAL.*Brain source localization using a physics-driven structured cosparse representation of EEG signals*, in "2014 IEEE International Workshop on Machine Learning for Signal Processing", Reims, France, September 2014, 6, https://hal.archives-ouvertes.fr/hal-01027609.

[57] S. ARBERET, A. OZEROV, F. BIMBOT, R. GRIBONVAL.*A tractable framework for estimating and combining spectral source models for audio source separation*, in "Signal Processing", August 2012, vol. 92, n^{o} 8, p. 1886-1901, http://hal.inria.fr/hal-00694071.

[58] C. BILEN, S. KITIĆ, N. BERTIN, R. GRIBONVAL.*Sparse Acoustic Source Localization with Blind Calibration for Unknown Medium Characteristics*, August 2014, iTwist - 2nd international - Traveling Workshop on Interactions between Sparse models and Technology, https://hal.inria.fr/hal-01060320.

[59] C. BLANDIN, A. OZEROV, E. VINCENT.*Multi-source TDOA estimation in reverberant audio using angular spectra and clustering*, in "Signal Processing", March 2012, vol. 92, p. 1950-1960, Revised version including minor corrections in equations (17), (18) and Figure 1 compared to the version published by Elsevier [*DOI :* 10.1016/J.SIGPRO.2011.10.032], http://hal.inria.fr/inria-00630994.

[60] A. BOURRIER.*Compressed sensing and dimensionality reduction for unsupervised learning*, Université Rennes 1, May 2014, https://tel.archives-ouvertes.fr/tel-01023030.

[61] A. BOURRIER, R. GRIBONVAL, P. PÉREZ.*Compressive Gaussian Mixture Estimation*, in "Signal Processing with Adaptive Sparse Structured Representations (SPARS) 2013", Switzerland, July 2013, http://hal.inria.fr/hal-00811819.

[62] A. BOURRIER, R. GRIBONVAL, P. PÉREZ.*Compressive Gaussian Mixture Estimation*, in "ICASSP - 38th International Conference on Acoustics, Speech, and Signal Processing", Vancouver, Canada, 2013, p. 6024-6028, http://hal.inria.fr/hal-00799896.

[63] A. BOURRIER, R. GRIBONVAL, P. PÉREZ.*Estimation de mélange de Gaussiennes sur données compressées*, in "24ème Colloque Gretsi", France, September 2013, 191, http://hal.inria.fr/hal-00839579.

[64] M. CHAFII.*Study of a new multicarrier waveform with low PAPR*, CentraleSupélec, October 2016, https://hal.archives-ouvertes.fr/tel-01399509.

[65] M. CHAFII, J. PALICOT, R. GRIBONVAL.*A PAPR upper bound of generalized waveforms for multi-carrier modulation systems*, in "6th International Symposium on Communications, Control, and Signal Processing - ISCCSP 2014", Athènes, Greece, May 2014, p. 461 - 464 [*DOI :* 10.1109/ISCCSP.2014.6877913], https://hal-supelec.archives-ouvertes.fr/hal-01072519.

[66] M. CHAFII, J. PALICOT, R. GRIBONVAL.*Closed-form Approximations of the PAPR Distribution for Multi-Carrier Modulation Systems*, in "EUSIPCO 2014 - European Signal Processing Conference", Lisbonne, Portugal, September 2014, https://hal.inria.fr/hal-01054126.

[67] M. CHAFII, J. PALICOT, R. GRIBONVAL.*Closed-form approximations of the peak-to-average power ratio distribution for multi-carrier modulation and their applications*, in "EURASIP Journal on Advances in Signal Processing", 2014, vol. 2014, n^o 1, 121 [*DOI :* 10.1186/1687-6180-2014-121], https://hal.inria.fr/hal-01056153.

[68] M. CHAFII, J. PALICOT, R. GRIBONVAL.*L'optimalité de l'OFDM en termes de performance en PAPR*, in "25ème Colloque Gretsi 2015", Lyon, France, September 2015, https://hal-supelec.archives-ouvertes.fr/hal-01165509.

[69] F. CHAUMETTE, S. HUTCHINSON.*Visual servo control, Part I: Basic approaches*, in "IEEE Robotics and Automation Magazine", 2006, vol. 13, n^o 4, p. 82-90, https://hal.inria.fr/inria-00350283.

[70] A. DELEFORGE, F. FORBES, R. HORAUD.*Acoustic space learning for sound-source separation and localization on binaural manifolds*, in "International journal of neural systems", 2015, vol. 25, n^o 01, 1440003, https://hal.archives-ouvertes.fr/hal-00960796.

[71] N. DUONG, E. VINCENT, R. GRIBONVAL.*Spatial location priors for Gaussian model-based reverberant audio source separation*, Inria, September 2012, n^o RR-8057, http://hal.inria.fr/hal-00727781.

[72] N. DUONG, E. VINCENT, R. GRIBONVAL.*Spatial location priors for Gaussian model based reverberant audio source separation*, in "EURASIP Journal on Advances in Signal Processing", September 2013, 149 [*DOI :* 10.1186/1687-6180-2013-149], http://hal.inria.fr/hal-00865125.

[73] R. GIRYES, S. NAM, M. ELAD, R. GRIBONVAL, M. E. DAVIES.*Greedy-Like Algorithms for the Cosparse Analysis Model*, in "Linear Algebra and its Applications", January 2014, vol. 441, p. 22–60, partially funded by the ERC, PLEASE project, ERC-2011-StG-277906 [*DOI :* 10.1016/J.LAA.2013.03.004], http://hal.inria.fr/hal-00716593.

[74] N. ITO.*Robust microphone array signal processing against diffuse noise*, University of Tokyo, January 2012, http://hal.inria.fr/tel-00691931.

[75] N. ITO, E. VINCENT, N. ONO, S. SAGAYAMA.*Robust estimation of directions-of-arrival in diffuse noise based on matrix-space sparsity*, Inria, October 2012, n^o RR-8120, http://hal.inria.fr/hal-00746271.

[76] N. KERIVEN, R. GRIBONVAL.*Compressive Gaussian Mixture Estimation by Orthogonal Matching Pursuit with Replacement*, July 2015, SPARS 2015, https://hal.inria.fr/hal-01165984.

[77] S. KITIĆ, N. BERTIN, R. GRIBONVAL.*A review of cosparse signal recovery methods applied to sound source localization*, in "Le XXIVe colloque Gretsi", Brest, France, September 2013, http://hal.inria.fr/hal-00838080.

[78] S. KITIĆ, N. BERTIN, R. GRIBONVAL.*Audio Declipping by Cosparse Hard Thresholding*, August 2014, iTwist - 2nd international - Traveling Workshop on Interactions between Sparse models and Technology, https://hal.inria.fr/hal-00922497.

[79] S. KITIĆ, N. BERTIN, R. GRIBONVAL.*Hearing behind walls: localizing sources in the room next door with cosparsity*, in "ICASSP - IEEE International Conference on Acoustics, Speech, and Signal Processing", Florence, Italy, May 2014 [*DOI :* 10.1109/ICASSP.2014.6854168], https://hal.inria.fr/hal-00904779.

[80] S. KITIĆ, N. BERTIN, R. GRIBONVAL.*Sparsity and cosparsity for audio declipping: a flexible non-convex approach*, in "LVA/ICA 2015 - The 12th International Conference on Latent Variable Analysis and Signal Separation", Liberec, Czech Republic, August 2015, 8, https://hal.inria.fr/hal-01159700.

[81] S. KITIĆ.*Cosparse regularization of physics-driven inverse problems*, Université Rennes 1 ; Inria Rennes Bretagne Atlantique, November 2015, https://hal.archives-ouvertes.fr/tel-01237323.

[82] M. KOWALSKI, K. SIEDENBURG, M. DÖRFLER.*Social Sparsity! Neighborhood Systems Enrich Structured Shrinkage Operators*, in "IEEE Transactions on Signal Processing", May 2013, vol. 61, n^{o} 10, p. 2498 - 2511 [*DOI :* 10.1109/TSP.2013.2250967], https://hal.archives-ouvertes.fr/hal-00691774.

[83] L. LE MAGOAROU, R. GRIBONVAL, A. GRAMFORT.*FAμST: speeding up linear transforms for tractable inverse problems*, in "European Signal Processing Conference (EUSIPCO)", Nice, France, August 2015, https://hal.archives-ouvertes.fr/hal-01156478.

[84] L. LE MAGOAROU, R. GRIBONVAL.*Chasing butterflies: In search of efficient dictionaries*, in "International Conference on Acoustics, Speech and Signal Processing (ICASSP)", Brisbane, Australia, April 2015 [*DOI :* 10.1109/ICASSP.2015.7178579], https://hal.archives-ouvertes.fr/hal-01104696.

[85] L. LE MAGOAROU.*Efficient matrices for signal processing and machine learning*, INSA de Rennes, November 2016, https://tel.archives-ouvertes.fr/tel-01412558.

[86] A. MAGASSOUBA, N. BERTIN, F. CHAUMETTE.*Sound-based control with two microphones*, in "IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, IROS'15", Hamburg, Germany, September 2015, p. 5568-5573, https://hal.inria.fr/hal-01185841.

[87] S. NAM, R. GRIBONVAL.*Physics-driven structured cosparse modeling for source localization*, in "Acoustics, Speech and Signal Processing, IEEE International Conference on (ICASSP 2012)", Kyoto, Japon, IEEE, 2012, http://hal.inria.fr/hal-00659405.

[88] A. A. NUGRAHA, A. LIUTKUS, E. VINCENT.*Multichannel music separation with deep neural networks*, in "24th European Signal Processing Conference (EUSIPCO 2016)", 2016, p. 1748–1752.

[89] A. OZEROV, C. FÉVOTTE.*Multichannel nonnegative matrix factorization in convolutive mixtures for audio source separation*, in "IEEE Transactions on Audio, Speech, and Language Processing", 2010, vol. 18, n^{o} 3, p. 550–563.

[90] G. PUY, M. E. DAVIES, R. GRIBONVAL.*Linear embeddings of low-dimensional subsets of a Hilbert space to $\mathbb{R}^m$*, in "EUSIPCO - 23rd European Signal Processing Conference", Nice, France, August 2015, https://hal.inria.fr/hal-01116153.

[91] G. PUY, M. E. DAVIES, R. GRIBONVAL.*Recipes for stable linear embeddings from Hilbert spaces to $\mathbb{R}^m$*, September 2015, Submitted to IEEE Transactions on Information Theory, https://hal.inria.fr/hal-01203614.

[92] Y. SALAÜN, E. VINCENT, N. BERTIN, N. SOUVIRAÀ-LABASTIE, X. JAUREGUIBERRY, D. TRAN, F. BIMBOT.*The Flexible Audio Source Separation Toolbox Version 2.0*, May 2014, ICASSP, https://hal.inria.fr/hal-00957412.

Project-Team SERPICO

Space-timE RePresentation, Imaging and cellular dynamics of molecular COmplexes

RESEARCH CENTER
Rennes - Bretagne-Atlantique

THEME
Computational Biology

Table of contents

Project-Team SERPICO

Creation of the Team: 2010 January 01, updated into Project-Team: 2013 July 01

Keywords:

Computer Science and Digital Science:

3.1.1. - Modeling, representation
3.1.2. - Data management, quering and storage
3.3. - Data and knowledge analysis
3.4. - Machine learning and statistics
5.3. - Image processing and analysis
5.3.2. - Sparse modeling and image representation
5.3.3. - Pattern recognition
5.3.4. - Registration
5.4.4. - 3D and spatio-temporal reconstruction
5.4.5. - Object tracking and motion analysis
5.4.6. - Object localization
5.9.1. - Sampling, acquisition
5.9.2. - Estimation, modeling
5.9.3. - Reconstruction, enhancement
5.9.6. - Optimization tools
6.1.2. - Stochastic Modeling (SPDE, SDE)
6.1.3. - Discrete Modeling (multi-agent, people centered)
6.1.4. - Multiscale modeling
6.1.5. - Multiphysics modeling
6.2.3. - Probabilistic methods
6.2.4. - Statistical methods
6.2.6. - Optimization
6.3.1. - Inverse problems
6.3.2. - Data assimilation
6.3.3. - Data processing

Other Research Topics and Application Domains:

1.1.1. - Structural biology
1.1.3. - Cellular biology
1.1.9. - Bioinformatics
1.1.10. - Mathematical biology
2.2.3. - Cancer
2.6. - Biological and medical imaging

1. Members

Research Scientists

Charles Kervrann [Team leader, Inria, Senior Researcher, HDR]

Patrick Bouthemy [Inria, Senior Researcher, HDR]
Sylvain Prima [Inria, Researcher, from Sept 2016]

Faculty Member

Frédéric Lavancier [Univ. Nantes, Faculty Member, Associate Professor ("Inria delegation"), from Sept 2014]

Technical Staff

Giovanni Petrazzuoli [Inria, from Jul 2016]
Ancageorgiana Caranfil [Inria, granted by ANR France-BioImaging project, from Dec 2015 to Nov 2016]
Thierry Pécot [Inria, granted by ANR France-BioImaging project, from Sept 2013 to Aug 2016]

PhD Students

Vincent Briane [Inria, Inria & ENSAI-CREST, from Oct 2014]
Anca-Georgiana Caranfil [Univ. Rennes I, from Dec 2015]
Sandeep Manandhar [Inria, from Oct 2016]
Emmanuel Moebel [Inria, granted by Région Bretagne, from Nov 2015]
Juan Manuel Perez Rua [Technicolor, granted by CIFRE]
Bertha Mayela Toledo Acosta [Conacyt Grant, from Oct 2014]
Hoai Nam Nguyen [Inria, granted by Innopsys Company, from Oct 2013 to Nov 2016]

Administrative Assistant

Huguette Bechu [Inria]

Other

Aminata Diouf [Inria, Internship Master 1 INSA-Rouen, from Jun 2016 until Aug 2016]

2. Overall Objectives

2.1. Glossary

FLIM (Fluorescence Lifetime Microscopy Imaging): imaging of fluorescent molecule lifetimes.

PALM (Photo-Activated Localization Microscopy): high-resolution microscopy using stochastic photo-activation of fluorophores and adjustment of point spread functions [41].

SIM (Structured Illumination Microscopy): high-resolution light microscopy using structured patterns and interference analysis [49].

TIRF (Total Internal Reflectance): 2D optical microscopy using evanescent waves and total reflectance [40].

Cryo-EM (Cryo Electron Tomography): 3D representation of sub-cellular and molecular objects of 5-20 nanometres, frozen at very low temperatures, from 2D projections using a transmission electron microscope.

2.2. Scientific context and motivations

Light microscopy, especially fluorescence microscopy, has taken a prominent role in life science research due to its ability to investigate the 3D interior of cells and organisms. It enables to visualize, in vitro and in vivo, particular biomolecules and proteins (gene expression) with high specificity through fluorescent labeling (GFP - Green Fluorescence Protein probes) both at the microscopic and nanoscopic scales. Nevertheless, the mechanisms of life are very complex and driven by multimolecular interactions: mitotic spindle, cell signaling complexes, intracellular transport, cell morphogenesis and motility... A dynamical quantitative and integrated description of molecular interactions and coordination within macromolecular complexes at different scales appears essential today for the global understanding of live mechanisms. A long-term research consists in inferring the relationships between the dynamics of macromolecules and their functions. This constitutes one of the challenges of modern biology. The proposed mathematical models and algorithms are mainly developed to identify molecular processes in fundamental biology but they have also a strong potential for applications in biotechnology and medicine: disease diagnosis, detection of genomic instabilities, deterioration of cell cycle, epigenetic mechanisms and cancer prevention.

2.3. Objectives in cell imaging

Facing the amount of information provided by high-throughput multidimensional microscopy, the SERPICO team investigates computational and statistical models to better elucidate the role of specific proteins inside their multiprotein complexes and to help to decipher the dynamic coordination and organization of molecular complexes at the single cell level. We investigate image processing methods, mathematical models, and algorithms to build an integrated imaging approach that bridges the resolution gaps between the molecule and the whole cell, in space and time [63]. We address the following topics:

- Image superresolution/image denoising required to preserve cell integrity (photo-toxicity versus exposure time) and image analysis in multidimensional microscopy;
- Motion analysis and computation of molecule trajectories in live-cell imaging to study molecular interactions in space and time);
- Computational simulation and modelling of molecule trafficking at different spatial and temporal scales (e.g. biophysical model assimilation for dynamic representation in video-microscopy and prediction in biology.

We focus on the cellular and molecular mechanisms involved in membrane transport and trafficking at the scale of a single cell.

2.4. Main challenges in image processing for multimodal and multidimensional microscopy

In most cases, modern microscopy in biology is characterized by a large number of dimensions that fits perfectly with the complexity of biological features: two or three spatial dimensions, at macro to nano-scales, and one temporal dimension, sometimes spectrally defined and often corresponding to one particular bio-molecular species. Dynamic microscopy is also characterized by the nature of the observable objects (cells, organelles, single molecules, ...), by the large number of small size and mobile elements (chromosomes, vesicles, ...), by the complexity of the dynamic processes involving many entities or group of entities sometimes interacting, by particular phenomena of coalescence often linked to image resolution problems, finally by the association, dissociation, recomposition or constitution of those entities (such as membrane fusion and budding). Thus, the corpus of data to be considered for a comparative analysis of multiple image series acquisitions is massive (up to few GigaBytes per hour). Therefore, it becomes necessary to facilitate and rationalize the production of those multidimensional data, to improve post acquisition analysis (i.e. image processing) which are limiting factors in front of the data, and to favor the organization and the interpretation of the information associated to this data corpus. It motivates and requires innovative mathematical tools and concepts: data fusion, image registration, superresolution, data mining, life dynamics modelling, ...

2.5. Organization and collaborations

In collaboration with UMR 144 CNRS-Institut Curie ("Space Time imaging of Endomembranes and organelles Dynamics" team) and PICT-IBiSA (Cell and Tissue Imaging Facilities), the members of the SERPICO team participate in several projects (PhD and post-doc supervision, contracts...) with biologists in the field of cell biology and microscopy. We have promoted and designed non-parametric methods since prior knowledge cannot be easily taken into account for extracting unattended but desired information from image data. We have proposed user-friendly algorithms for processing 3D or 4D data.

The scientific projects of the SERPICO team are complementary to the other on-going and planned projects of the UMR 144 CNRS-Institut Curie Unit. A subset of projects is related to instrumentation in electronic and photonic microscopy (PICT-IBiSA platform) including computational aspects on the reconstruction and enhancement of images related to sub-diffraction light microscopy and multimodal approaches. Our projects rely partially on the results and advances of these instrumental projects and a positive synergy is foreseen.

3. Research Program

3.1. Statistics and algorithms for computational microscopy

Many live-cell fluorescence imaging experiments are limited in time to prevent phototoxicity and photobleaching. The amount of light and time required to observe entire cell divisions can generate biological artifacts. In order to produce images compatible with the dynamic processes in living cells as seen in video-microscopy, we study the potential of denoising, superresolution, tracking, and motion analysis methods in the Bayesian and the robust statistics framework to extract information and to improve image resolution while preserving cell integrity.

In this area, we have already demonstrated that image denoising allows images to be taken more frequently or over a longer period of time [6]. The major advantage is to preserve cell integrity over time since spatio-temporal information can be restored using computational methods [9], [3], [10], [5]. This idea has been successfully applied to wide-field, spinning-disk confocal microscopy [2], TIRF [40], fast live imaging and 3D-PALM using the OMX system in collaboration with J. Sedat and M. Gustafsson at UCSF [6]. The corresponding ND-SAFIR denoiser software (see Section 6.7) has been licensed to a large set of laboratories over the world. New information restoration and image denoising methods are currently investigated to make SIM imaging compatible with the imaging of molecular dynamics in live cells. Unlike other optical sub-diffraction limited techniques (e.g. STED [51], PALM [41]) SIM has the strong advantage of versatility when considering the photo-physical properties of the fluorescent probes [49]. Such developments are also required to be compatible with "high-throughput microscopy" since several hundreds of cells are observed at the same time and the exposure times are typically reduced.

3.2. From image data to descriptors: dynamic analysis and trajectory computation

3.2.1. Motion analysis and tracking

The main challenge is to detect and track xFP tags with high precision in movies representing several Giga-Bytes of image data. The data are most often collected and processed automatically to generate information on partial or complete trajectories. Accordingly, we address both the methodological and computational issues involved in object detection and multiple objects tracking in order to better quantify motion in cell biology. Classical tracking methods have limitations as the number of objects and clutter increase. It is necessary to correctly associate measurements with tracked objects, i.e. to solve the difficult data association problem [57]. Data association even combined with sophisticated particle filtering techniques [60] or matching techniques [58] is problematic when tracking several hundreds of similar objects with variable velocities. Developing new optical flow and robust tracking methods and models in this area is then very stimulating since the problems we have to solve are really challenging and new for applied mathematics. In motion analysis, the goal is to formulate the problem of optical flow estimations in ways that take physical causes of brightness constancy violations into account [46], [50]. The interpretation of computed flow fields enables to provide spatio-temporal signatures of particular dynamic processes (e.g. Brownian and directed motion) and could help to complete the traffic modelling.

3.2.2. Event detection and motion classification

Protein complexes in living cells undergo multiple states of local concentration or dissociation, sometimes associated with diffusion processes. These events can be observed at the plasma membrane with TIRF microscopy. The difficulty arises when it becomes necessary to distinguish continuous motions due to trafficking from sudden events due to molecule concentrations or their dissociations. Typically, plasma membrane vesicle docking, membrane coat constitution or vesicle endocytosis are related to these issues.

Several approaches can be considered for the automatic detection of appearing and vanishing particles (or spots) in wide-field and TIRF microscopy images. Ideally this could be performed by tracking all the vesicles contained in the cell [60], [48]. Among the methods proposed to detect particles in microscopy images [61], [59], none is dedicated to the detection of a small number of particles appearing or disappearing suddenly between two time steps. Our way of handling small blob appearances/dis-appearances originates from the observation that two successive images are redundant and that occlusions correspond to blobs in one image which cannot be reconstructed from the other image [2] (see also [44]). Furthermore, recognizing dynamic protein behaviors in live cell fluorescence microscopy is of paramount importance to understand cell mechanisms. In our studies, it is challenging to classify intermingled dynamics of vesicular movements, docking/tethering, and ultimately, plasma membrane fusion of vesicles that leads to membrane diffusion or exocytosis of cargo proteins. Our aim is then to model, detect, estimate and classify subcellular dynamic events in TIRF microscopy image sequences. We investigate methods that exploits space-time information extracted from a couple of successive images to classify several types of motion (directed, diffusive (or Brownian) and confined motion) or compound motion.

3.3. From models to image data: simulation and modelling of membrane transport

Mathematical biology is a field in expansion, which has evolved into various branches and paradigms to address problems at various scales ranging from ecology to molecular structures. Nowadays, system biology [52], [63] aims at modelling systems as a whole in an integrative perspective instead of focusing on independent biophysical processes. One of the goals of these approaches is the cell in silico as investigated at Harvard Medical School (http://vcp.med.harvard.edu/) or the VCell of the University of Connecticut Health Center (http://www.nrcam.uchc.edu/). Previous simulation-based methods have been investigated to explain the spatial organization of microtubules [53] but the method is not integrative and a single scale is used to describe the visual patterns. In this line of work, we propose several contributions to combine imaging, traffic and membrane transport modelling in cell biology.

In this area, we focus on the analysis of transport intermediates (vesicles) that deliver cellular components to appropriate places within cells. We have already investigated the concept of Network Tomography (NT) [62] mainly developed for internet traffic estimation. The idea is to determine mean traffic intensities based on statistics accumulated over a period of time. The measurements are usually the number of vesicles detected at each destination region receiver. The NT concept has been investigated also for simulation [4] since it can be used to statistically mimic the contents of real traffic image sequences. In the future, we plan to incorporate more prior knowledge on dynamics to improve representation. An important challenge is to correlate stochastic, dynamical, one-dimensional *in silico* models studied at the nano-scale in biophysics, to 3D images acquired in vivo at the scale of few hundred nanometers.

4. Application Domains

4.1. Modeling and analysis of membrane transport and molecule trafficking at the single cell scale

In the past recent years, research carried at UMR 144 CNRS-Institut Curie ("Space Time imaging of Endomembranes and organelles Dynamics" team) contributed to a better understanding of the intracellular compartimentation of specialized model cells such as melanocytes and Langerhans cells, the components and structural events involved in the biogenesis of their specialized organelles: melanosomes and Birbeck granules, respectively. These studies have started to highlight: i/ multiple sorting and structural events involved in the biogenesis of these organelles; ii/ complexity of the endo-melanosomal network of these highly specialized cells; iii/ complex molecular architecture organizing and coordinating their dynamics; iv/ intracellular transport steps affected in genetic diseases, among which the Hermansky Pudlak syndrome (HPS) or involved in viral infection (HIV and Langerin in Langerhans cells).

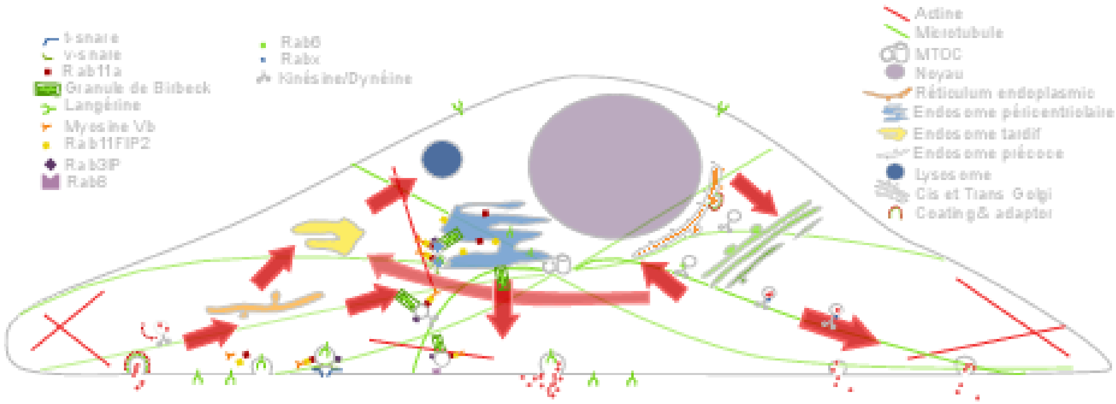


Figure 1. Cargo Langerin Trafficking controlled by Rab11A/Rab11FIP2/MyoVb platform.

In this context, the central aim of SERPICO is to understand how the different machineries of molecular components involved are interconnected and coordinated to generate such specialized structures. We need to address the following topics:

1. developing new bioimaging approaches to observe and statistically analyze such coordinated dynamics in live material;
2. correlating this statistically relevant spatiotemporal organization of protein networks with the biological architectures and at the ultrastructural level;
3. modeling intracellular transport of those reference biological complex systems and proposing new experimental plans in an iterative and virtuous circle;
4. managing and analyzing the workflow of image data obtained along different multidimensional microscopy modalities.

These studies are essential to unravel the complexity of the endomembrane system and how different machineries evolve together (e.g. see Fig. 1). They help to control cell organization and function at different scales through an integrative workflow of methodological and technological developments.

At long term, these studies will shed light on the cellular and molecular mechanisms underlying antigen presentation, viral infection or defense mechanisms, skin pigmentation, the pathogenesis of hereditary genetic disorders (lysosomal diseases, immune disorders) and on the mechanisms underlying cell transformation. Our methodological goal is also to link dynamics information obtained through diffraction limited light microscopy, eventually at a time regime compatible with live cell imaging. The overview of ultrastructural organization will be achieved by complementary electron microscopical methods. Image visualization and quantitative analysis are of course important and essential issues in this context.

4.2. Imaging and analysis of cytokskeleton dynamics during cell migration

The ability to migrate in space is among the most fundamental functions of eukaryotic cells and thus is one of the best-studied phenomena in biology. During embryonic development, cell movements result in a massive reorganization of the embryo, from a simple spherical ball of cells into a multi-layered organism; many of the cells at or near the surface of the embryo move to a new, more interior location. Moreover, inadequate or inappropriate migration of immune cells is also critically important for the delivery of protective immune responses to tissues and for wound healing. Finally, cell migration may facilitate the dissemination of tumor cells in blood and organs and eventually the formation of secondary tumors and metastases.

It has been established that the cytoskeleton, composed of actin filaments, microtubules and intermediate filaments (elongated structures with a diameter of a few dozens of nanometers), is essential for several cell mechanisms, including cell migration, cell division and molecule trafficking:

- i/ the actin filaments promote cell protrusion, adhesion and retraction;
- ii/ the microtubules are the support of molecule traffic and cell polarization;
- iii/ the intermediate filaments are hypothesized to control microtubule organization.

Nevertheless, the mechanical and chemical states of migrating cells under various external conditions remain largely unknown. In the last decade, high-resolution microscopy methods led to the discovery of novel aspects of cell migration. Most approaches and models are limited to migration in 2D, justified by the flatness of the cell-motile mechanisms. However, the mechanical patterns that govern migration in 2D models are often not essential for efficient migration in 3D. Accordingly, recent very challenging 3D models of cells moving on flat surfaces have begun to emerge. The key challenge, however, is to understand how a 3D motile cell crawls through the 3D extracellular matrix.

The objective of SERPICO is to develop high-end signal processing and computer vision tools to unfold the dynamical coordination of microtubules, actin filaments and intermediate filaments in 3D, involved in cell migration, cell division and molecule trafficking.

5. Highlights of the Year

5.1. Highlights of the Year

5.1.1. New projects

The 4 year-ANR-DALLISH proposal (PRC / Challenge 7 / Topic 5), coordinated by the Serpico Team-Project, has been accepted in September 2016.

The CytoDI Associated Team, in collaboration with University of Texas, SouthWestern Medical Center, Dallas (TX, USA) started in January 2016.

6. New Software and Platforms

6.1. TMA-Lib

KEYWORD: Biomedical imaging
FUNCTIONAL DESCRIPTION: The TMA-LIB enables to jointly detect (adaptive wavelet transform), segment (parametric active contours) and restore (artifact correction and deconvolution) TMA (Tissue MicroArrays) images.

- **Participants**: Hoai Nam Nguyen, Charles Kervrann.
- **Partner:** INNOPSYS Company.
- **Contact:** Charles Kervrann.
- **Languages:** C/C++, Matlab.

6.2. QuantEv

KEYWORD: Biomedical imaging

FUNCTIONAL DESCRIPTION: The QUANTEV software is dedicated to the analysis of the spatial distribution of intracellular events represented by any static or dynamical descriptor (e.g. detected points, segmented regions, trajectories...), provided that the descriptors are associated with spatial coordinates. QuantEv first computes 3D histograms of descriptors in a cylindrical coordinate system with computational cell shape normalization, enabling comparisons between cells of different shape. Densities are obtained via adaptive kernel density estimation, and we use the Circular Earth Mover's Distance to measure the dissimilarity between densities associated to different experimental conditions. A statistical analysis on these distances reliably takes into account the biological variability over replicated experiments.

- **Participants**: Thierry Pécot, Charles Kervrann, Jean Salamero.
- **Contact:** Thierry Pécot, Charles Kervrann.
- **On-line demo:** http://mobyle-serpico.rennes.inria.fr/cgi-bin/portal.py#forms::...
- **Languages:** C/C++ and JAVA (plug-in ICY: http://icy.bioimageanalysis.org).

6.3. C-CRAFT

KEYWORD: Biomedical imaging
FUNCTIONAL DESCRIPTION: The C-CRAFT software enables to jointly segment small particles and estimate background in 2D or 3D fluorescence microscopy image sequences. The vesicle segmentation and background estimation problem is formulated as a global energy minimization problem in the Conditional Random Field framework. A patch-based image representation is used to detect spatial irregularity in the image. An iterative scheme based on graph-cut algorithm is proposed for energy minimization.

- **Participants**: Thierry Pécot, Charles Kervrann, Patrick Bouthemy, Jean Salamero.
- **Contact:** Thierry Pécot, Charles Kervrann.
- **On-line demo:** http://mobyle-serpico.rennes.inria.fr/cgi-bin/portal.py#forms::C-CRAFT
- **Languages:** C/C++ and JAVA (plug-in ICY: http://icy.bioimageanalysis.org/plugin/C-CRAFT).
- **Reference:** [12]

6.4. ATLAS

KEYWORD: Biomedical imaging
FUNCTIONAL DESCRIPTION: The ATLAS software enables to detect spots in 2D fluorescence images. The spot size is automatically selected and the detection threshold adapts to the local image dynamics. ATLAS relies on the Laplacian of Gaussian (LoG) filter, which both reduces noise and enhances spots. A multiscale representation of the image is built to automatically select the optimal LoG variance. Local statistics of the LoG image are estimated in a Gaussian window, and the detection threshold is pointwise inferred from a probability of false alarm (PFA). The user only has to specify: i/ standard deviation of the Gaussian window; ii/ PFA value. The Gaussian window must be about the size of the background structures; increasing the PFA increases the number of detections.

- **Participants**: Antoine Basset, Patrick Bouthemy, Thierry Pécot, Charles Kervrann.
- **Contact:** Thierry Pécot, Patrick Bouthemy, Charles Kervrann.
- **On-line demo:** http://mobyle-serpico.rennes.inria.fr/cgi-bin/portal.py#forms::ATLAS
- **Language:** C/C++.
- **Reference:** [1]

6.5. Hullkground

KEYWORDS: Bioinformatics - Biomedical imaging

FUNCTIONAL DESCRIPTION: The HULLKGROUND software decomposes a fluorescence microscopy image sequence into two dynamic components: i/ an image sequence showing mobile objects, ii/ an image sequence showing the slightly moving background. Each temporal signal of the sequence is processed individually and analyzed with computational geometry tools. The convex hull is estimated automatically for each pixel and subtracted to the original signal. The method is unsupervised, requires no parameter tuning and is a simplified version of the shape-based scale-space method.

- **Participants:** Anatole Chessel, Jean Salamero, Charles Kervrann.
- **Contact:** Charles Kervrann.
- **APP deposit number:** IDDN.FR.001.400005.000.S.P.2009.000.21000
- **On-line demo:** http://mobyle-serpico.rennes.inria.fr/cgi-bin/portal.py#forms::Hullkground
- **Free distribution:** http://serpico.rennes.inria.fr/doku.php?id=software:hullkground:hullkground
- **Language:** JAVA (plug-in IMAGEJ: http://rsbweb.nih.gov/ij/).

6.6. Motion2D

KEYWORDS: Image sequence - Motion model - 2D
FUNCTIONAL DESCRIPTION: The MOTION2D software is a multi-platform object-oriented library to estimate 2D parametric motion models in an image sequence. It can handle several types of motion models, namely, constant (translation), affine, and quadratic models. Moreover, it includes the possibility of accounting for a global variation of illumination and more recently for temporal image intensity decay (e.g. due to photobleaching decay in fluorescence microscopy). The use of such motion models has been proved adequate and efficient for solving problems such as optic flow computation, motion segmentation, detection of independent moving objects, object tracking, or camera motion estimation, and in numerous application domains (video surveillance, visual servoing for robots, video coding, video indexing), including biological imaging (image stack registration, motion compensation in videomicroscopy). Motion2D is an extended and optimized implementation of the robust, multi-resolution and incremental estimation method (exploiting only the spatio-temporal derivatives of the image intensity function). Real-time processing is achievable for motion models involving up to six parameters. Motion2D can be applied to the entire image or to any pre-defined window or region in the image.

- **Participants:** Patrick Bouthemy, Jean-Marc Odobez, Fabien Spindler.
- **Contact:** Patrick Bouthemy, Fabien Spindler.
- **APP deposit number:** FR.001.520021.001.S.A.1998.000.21000 / release 1.3.11, January 2005)
- **Free academic software distribution:** http://www.irisa.fr/vista/Motion2D
- **On-line demo:** http://mobyle-serpico.rennes.inria.fr/cgi-bin/portal.py#forms::Motion2D
- **Languages:** C/C++ and JAVA (plug-in IMAGEJ: http://rsbweb.nih.gov/ij/).

6.7. ND-SAFIR

KEYWORDS: Biology - Health - Image analysis - Photonic imaging - Fluorescence microscopy - Biomedical imaging

SCIENTIFIC DESCRIPTION: The ND-SAFIR software removes additive Gaussian and non-Gaussian noise in still 2D or 3D images or in 2D or 3D image sequences (without any motion computation) [5]. The method is unsupervised and is based on a pointwise selection of small image patches of fixed size (a data-driven adapted way) in spatial or space-time neighbourhood of each pixel (or voxel). The main idea is to modify each pixel (or voxel) using the weighted sum of intensities within an adaptive 2D or 3D (or 2D or 3D + time) neighbourhood and to use image patches to take into account complex spatial interactions. The neighbourhood size is selected at each spatial or space-time position according to a bias-variance criterion. The algorithm requires no tuning of control parameters (already calibrated with statistical arguments) and no library of image patches. The method has been applied to real noisy images (old photographs, JPEG-coded images, videos, ...) and is exploited in different biomedical application domains (time-lapse fluorescence microscopy, video-microscopy, MRI imagery, X-ray imagery, ultrasound imagery, ...).

- **Participants:** Jérôme Boulanger, Charles Kervrann, Patrick Bouthemy, Jean Salamero.
- **Partners:** INRA, PiCT - CNRS - Institut Curie.
- **APP deposit number:** IDDN.FR.001.190033.002.S.A.2007.000.21000 / new release 3.0 in 2013)
- **Free academic software distribution:** Binaries of the software ND-SAFIR are freely and electronically distributed (http://serpico.rennes.inria.fr/doku.php?id=software:nd-safir:index).
- **On-line demo:** http://mobyle-serpico.rennes.inria.fr/cgi-bin/portal.py#forms::NDSafir
- **Languages:** C/C++, MATLAB and JAVA (plug-in IMAGEJ: http://rsbweb.nih.gov/ij/). The C/C++ software has been developed under Linux using the CImg library and has been tested over several platforms such as Linux/Unix, Windows XP and Mac OS.
- **Commercial licence agreements:** Innopsys, Roper Scientfic, Photmetrics, Nikon Europe BV (2016).
- **Reference:** [5]

6.8. F2D-SAFIR

KEYWORD: Biomedical imaging
FUNCTIONAL DESCRIPTION: The F2D -SAFIR software removes mixed Gaussian-Poisson noise in large 2D images, typically $10^3 \times 10^3$ pixels, in a few seconds. The method is unsupervised and is a simplified version of the method related to the ND-SAFIR software. The software dedicated to microarrays image denoising, was licensed to the INNOPSYS company which develops scanners for disease diagnosis and multiple applications (gene expression, genotyping, aCGH, ChIP-chip, microRNA, ...).

- **Participant**: Charles Kervrann.
- **Partner:** INRA.
- **Contact:** Charles Kervrann.
- **APP deposit number:** IDDN.FR.001.190033.001.S.A.2007.000.21000
- **Language:** C/C++.

6.9. TubuleJ

KEYWORDS: Bioinformatics - Biomedical imaging

FUNCTIONAL DESCRIPTION: The TUBULEJ software written in java (plug-in ImageJ) is devoted to the analysis of microtubules and helical structures in 2D cryo electron microscope images. The software straightens curved microtubule images by estimating automatically point locations on the microtubule axis. The estimation of microtubule principal axis relies on microtubule cylindrical shape analyzed in the Fourier domain. A user-friendly interface enables to filter straight fiber images by selecting manually the layer lines of interest in the Fourier domain. This software can be used to generate a set of 2D projection views from a single microtubule projection view and a few parameters of this microtubule structure.

- **Participants:** Denis Chrétien, Charles Kervrann, Sophie Blestel.
- **Contact:** Denis Chrétien.
- **Partners:** University of Rennes 1, CNRS.
- **APP deposit number:** IDDN.FR.001.240023.000.S.P.2011.000.21000
- **On-line demo:** http://equipes.igdr.univ-rennes1.fr/en/tips/Software/TubuleJ/
- **Language:** JAVA (plug-in IMAGEJ: http://rsbweb.nih.gov/ij/).

6.10. Cryo-Seg

KEYWORDS: Bioinformatics - Biomedical imaging
FUNCTIONAL DESCRIPTION: The CRYO-SEG software has been developed to detect microtubule structures and helical structures in 2D cryo electron microscope images. Cryo electron tomography allows 3D observation of biological specimens in their hydrated state. Segmentation is formulated as Maximum A Posteriori estimation problem and exploits image patches to take into account spatial contexts (Markov Random Fields). Because of the contrast anisotropy in the specimen thickness direction, the whole tomogram is segmented section by section, with an automatic update of reference patches. This algorithm has been evaluated on synthetic data and on cryo electron tomograms of in vitro microtubules. On real data, this segmentation method extracts the most contrasted regions of microtubules, and 3D visualization is improved.

- **Participants:** Denis Chrétien, Charles Kervrann, Sophie Blestel.
- **Contact:** Denis Chrétien.
- **Partners:** University of Rennes 1, CNRS.
- **Languages:** C/C++ and JAVA (plug-in IMAGEJ: http://rsbweb.nih.gov/ij/).

6.11. Platforms

6.11.1. Mobyle@Serpico plateform and software distribution

The objective is to disseminate the distribution of SERPICO image processing software for biologists:

Free binaries: software packages have been compiled for the main operating systems (Linux, MacOS, Windows) using CMake (see http://www.cmake.org/). They are freely available on the team website under a proprietary license (e.g. ND-SAFIR and HULLKGROUND are distributed this way at http://serpico.rennes.inria.fr/doku.php?id=software:index).

Mobyle@SERPICO *web portal*: An on-line version of the image processing algorithms has been developped using the Mobyle framework (Institut Pasteur, see http://mobyle.pasteur.fr/). The main role of this web portal (see Fig. 2) is to demonstrate the performance of the programs developed by the team: QUANTEV, C-CRAFT[12], ATLAS[1], HOTSPOTDETECTION[56], HULLKGROUND[45], KLTRACKER[55], MOTION2D[54], MS-DETECT[47], ND-SAFIR[5], OPTICALFLOW and FLUX ESTIMATION [12]. The web interface makes our image processing methods available for biologists at Mobyle@SERPICO (http://mobyle-serpico.rennes.inria.fr/cgi-bin/portal.py#welcome) without any installation or configuration on their own. The size of submitted images is limited to 200 MegaBytes per user and all the results are kept 15 days. The web portal and calculations run on a server with 2 CPU x 8 cores, 64 GigaBytes of RAM (500 MegaBytes for each user / Data is saved for 3 months).

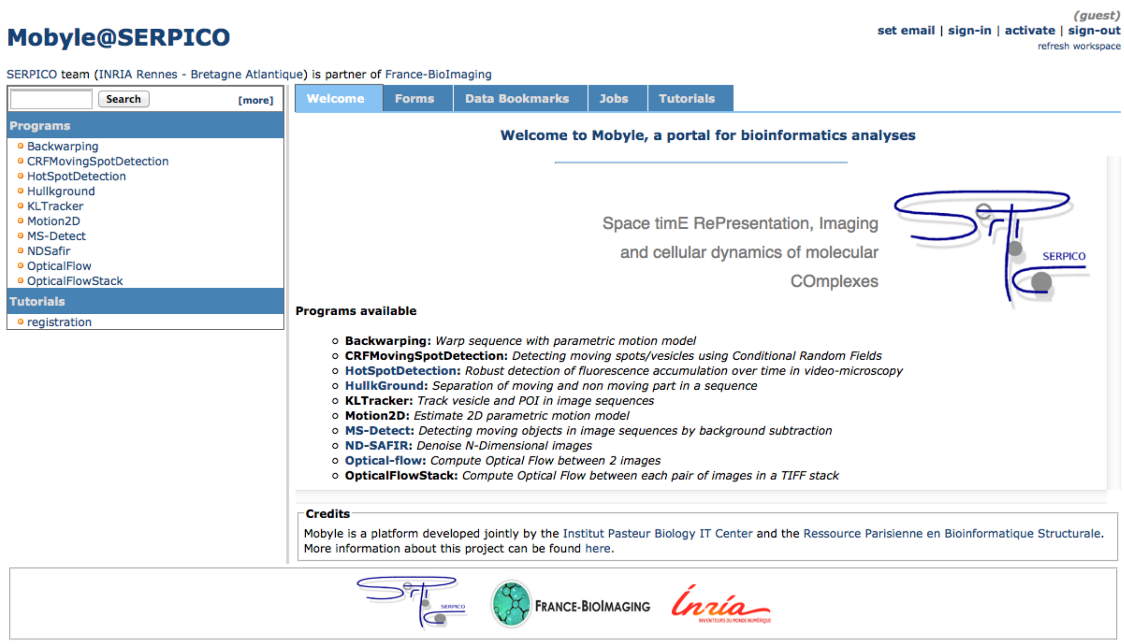


Figure 2. Mobyle@SERPICO web portal.

IMAGEJ **plug-ins**: IMAGEJ (see http://rsb.info.nih.gov/ij/) is a widely used image visualization and analysis software for biologist users. We have developed IMAGEJ plug-in JAVA versions of the following software: ND-SAFIR [5], HULLKGROUND [45], MOTION2D [54], HOTSPOTDETECTION [56], STLAS [1]. The C-CRAFT algorithm [12] has been developped for the image processing ICY platform (http://icy.bioimageanalysis.org/).

Institut Curie CID iManage database: The microscopy facility of Institut Curie has co-developped a commercial database system (CID iManage/Strand Avadis company). The database can be searched via meta-data and includes menu selections that enable to run remote processing from a cluster. We have integrated ND-SAFIR and HULLKGROUND in the interface environment to allow the database users to process their images easily, and store associated results and parameters used.

- **Participants:** Thierry Pécot, Charles Kervrann, Charles Deltel (Inria Rennes SED).
- **Contact:** Thierry Pécot, Charles Kervrann.

6.11.2. IGRIDA-Serpico cluster

The IGRIDA-Serpico cluster of 200 nodes is opened for end-users for large scale computing and data sets processing (200 TeraBytes).

- **Batch Scheduler:** OAR
- **File management:** Puppet / Git / Capistrano
- **OS:** Linux Debian 7
- **User connexion:** public ssh key
- **Contact:** Thierry Pécot, Charles Kervrann, Charles Deltel (Inria Rennes SED).

7. New Results

7.1. Statistical aggregation methods for image denoising and estimation

Participants: Charles Kervrann, Frédéric Lavancier.

In the line of the Non-Local means [43] and ND-SAFIR [9], [10], [5] denoising algorithms, we have proposed a novel adaptive estimator based on the weighted average of observations taken in a neighborhood with weights depending on image data. The idea is to compute adaptive weights that best minimize an upper bound of the pointwise L_2 risk. In the framework of adaptive estimation, we show that the "oracle" weights depend on the unknown image and are optimal if we consider triangular kernels instead of the commonly-used Gaussian kernel. Furthermore, we propose a way to automatically choose the spatially varying smoothing parameter for adaptive denoising. Under conventional minimal regularity conditions, the obtained estimator converges at the usual optimal rate. The implementation of the proposed algorithm is also straightforward. The simulations show that our algorithm improves significantly the classical NL-means, and is competitive when compared to the more sophisticated NL-means filters both in terms of PSNR values and visual quality.

Previously, we investigated statistical aggregation methods which optimally combine several estimators to produce a boosted solution [11]. In this range of work, we also introduced in [24] 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. This method has been subsequently adapted in [39] to models in spatial statistics. As part of an on-going work, we are applying this method to improve patch-based image denoising algorithms.

References: [24] [39]

Collaborators: Qiyu Jin (School of Mathematical Science, Inner Mongolia University, China),
Ion Grama and Quansheng Liu (University of Bretagne-Sud, Vannes),
Paul Rochet (Laboratoire de Mathématiques Jean Leray (LMJL), University of Nantes).

7.2. Algorithms for dejittering and deconvolving large fluorescence and Tissue MicroArray (TMA) images

Participants: Hoai Nam Nguyen, Giovanni Petrazzuoli, Aminata Diouf, Charles Kervrann.

In fluorescence microscopy, the image quality is limited by out-of-focus blur and high noise. Traditionally, image deconvolution is needed to estimate a good quality version of the observed image. The result of deconvolution depends heavily on the choice of the regularization term and the noise dependent fidelity term. The regularization functional should be designed to remove noise while preserving image discontinuities. Accordingly, we investigated new regularization terms to preserve fine details of underlying structures and we studied appropriate proximal algorithms. The deconvolution method has been especially dedicated to large 2D 20000×60000 images acquired with ISO scan imager (see Fig.3). The images are preliminary pre-processed to compensate non constant pixel displacement during acquisition/scanning (dejittering effect). The method has also been evaluated on 2D Vimentin filament images (UTSW, CytoDI Associated Team) to facilitate filament segmentation. The method is able to process a 512×512 image in 250 ms with a non optimized implementation.

Collaborators: Vincent Paveau and Cyril Cauchois (Innopys company),
Philippe Roudot (UTSW, Dallas, USA).

7.3. Quantifying the spatial distribution of intracellular events

Participants: Thierry Pécot, Charles Kervrann.

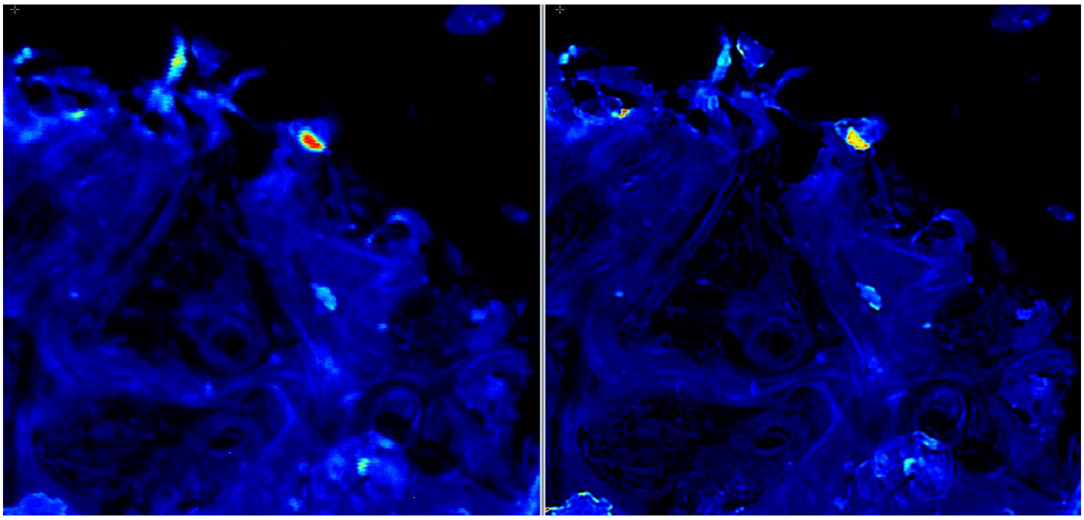


Figure 3. Illustration of image dejittering and deconvolution algorithms applied to a selected fluorescence region extracted from a large TMA image (by courtesy of Innopsys).

Automated processing of fluorescence microscopy data allows quantifying cell phenotypes in an objective and reproducible way. However, most computational methods are based on the complex combination of heterogeneous features such as statistical, geometrical, morphological and frequency properties, which makes difficult to draw definitive biological conclusions. Additionally, most experimental designs, especially at single cell level, pool together data coming from replicated experiments of a given condition, neglecting the biological variability between individual cells. To address these issues, we developed a generic and non-parametric framework (QuantEv) to study the spatio-temporal distribution of moving Rab6 membranes and the effect of actin disruption on Rab11 trafficking in coordination with cell shape. The main advantage of QuantEv is to process robustly and accurately homogeneous and heterogeneous populations. As demonstration, we compared the results obtained by QuantEv with those from kernel density maps, for Rab6 positive membranes on crossbow- and disk-shaped cells.

Collaborators: Jean Salamero, Jérôme Boulanger and Liu Zengzhen (UMR 144 CNRS-Institut Curie).

7.4. Correlation-based method for membrane diffusion estimation during exocytosis in TIRFM

Participants: Ancageorgiana Caranfil, Charles Kervrann.

The dynamics of the plasma membrane of the cell is not fully understood yet; one of the crucial aspects to clarify is the diffusion process during exocytosis. Several image acquisition modalities exist, including TIRFM (Total Internal Reflection Fluorescence Microscopy), that have successfully been used to determine the successive steps of exocytosis. However, computing characteristic values for plasma membrane dynamics is problematic, as the experimental conditions have a strong influence on the obtained data, and a general model of molecular interaction dynamics cannot be determined.

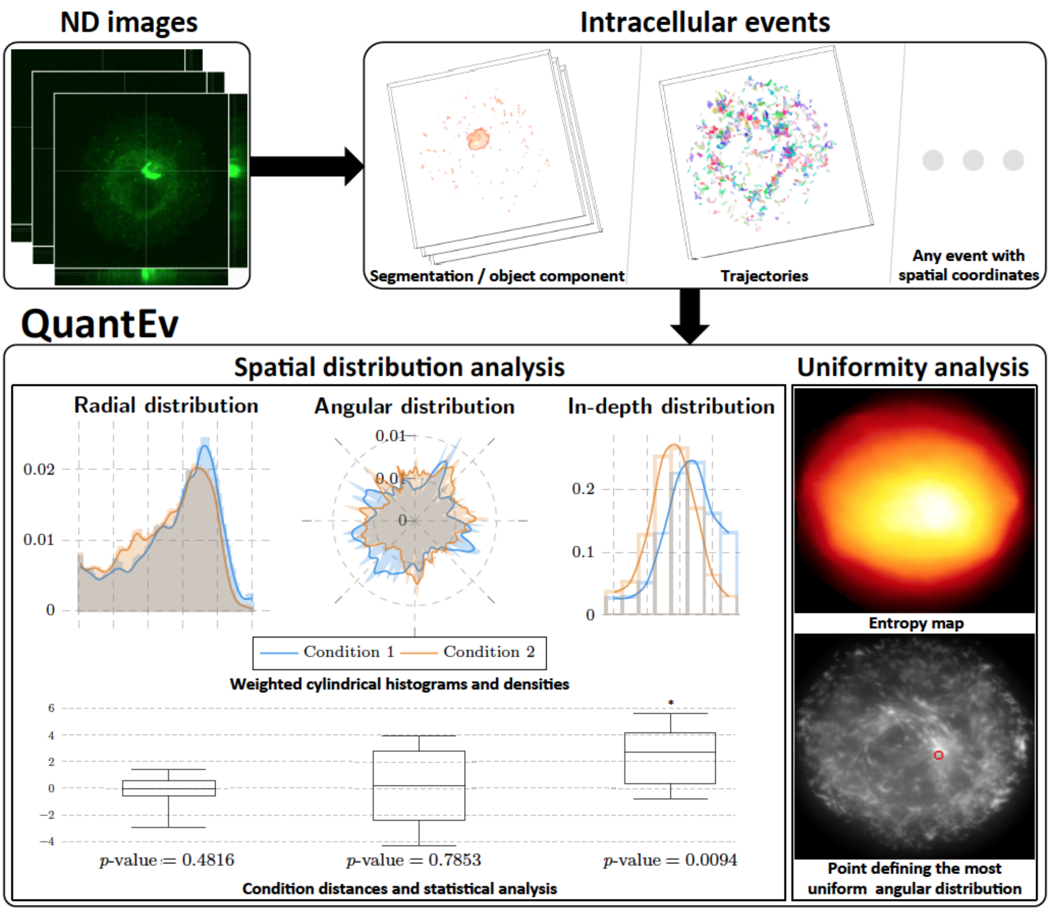


Figure 4. Overview of QuantEv approach.

This year, we have continued our study of correlation-based methods for local diffusion estimation in TIRFM images. Our original method was tested on both synthetic and real images showing an isolated diffusion event, and a robust algorithm was developed to cope with noisy data. Our first model was linear and had only two parameters to estimate. Diffusion coefficient estimation was accurate on synthetic images even with moderate to low signal-to-noise ratio, and within reasonable margins of error on real images with little noise. We have then extended our mathematical model by using a global approach subject to initial local diffusion conditions. Isolated diffusion events are well described, but this new model can also handle the case of noisy images with non-uniform background, and the case of two or more diffusion events in the region of interest. The extended model is non-linear but has few parameters to estimate. An iterative minimization method is used to fit the model parameters to the data points (see Fig. 5). Despite non-linearity, results are accurate on images with pure diffusion events and show robustness to background. The quality of parameters estimation is barely influenced by the length and size of the input TIRFM sequence, which is not the case with standard correlation methods. We have thus developed a correlation-based method that is able to estimate diffusion in a variety of cases in TIRFM images (Fig. 5).

Collaborators: Francois Waharte (UMR 144 CNRS-Institut Curie, PICT-IBiSA),
Perrine Paul-Gilloteaux (UMS 3556, IRS-UN, Nantes).

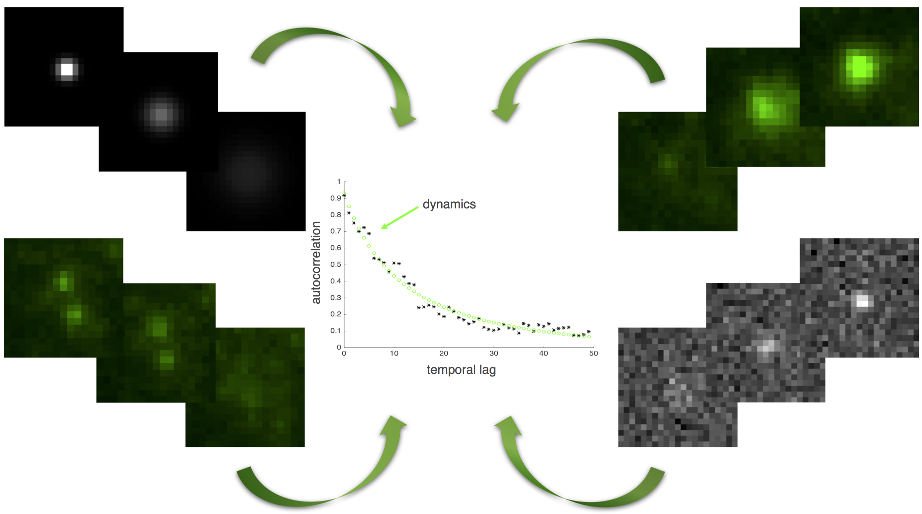


Figure 5. Local diffusion estimation in both synthetic (black and white spots) and real TIRF images (green spots, UMR 144 CNRS-Institut Curie, PICT-IBiSA). Only three frames of each sequence are shown. The computed autocorrelation values (in black) for different temporal lag values, and the fitting function for these values (in green) are given (plot in the middle) for the upper-right sequence.

7.5. Colocalization for live cell and super-resolution imaging

Participants: Frédéric Lavancier, Thierry Pécot, Charles Kervrann.

In the context of bioimaging, colocalization refers to the detection of emissions from two or more fluorescent molecules within the same pixel of the image. This approach enables to quantify the protein-protein interactions inside the cell, just at the resolution limit of the microscope. In statistics, this amounts to characterizing the joint spatial repartition and the spatial overlap between different fluorescent labels. Two distinct categories of colocalization approaches are considered to address this issue: intensity-based methods and object-based methods. The popular (intensity-based) Pearson's correlation method, which returns values between -1 and +1, is known to be sensitive to high intensity backgrounds and provides errors if the signal-to-noise ratio (SNR) is typically low. The object-based method, recommended in single molecule imaging, analyses the spatial distribution of the two sets of detected spots by using point process statistics.

Accordingly, we developed an original, fast, robust-to-noise and versatile approach that reconciles intensity-based and object-based methods for both conventional diffraction-limited microscopy and sub-resolved microscopy. The procedure is only controlled by a p-value and tests whether the Pearson correlation between two binary images is significantly positive. This amount to quantifying the interaction strength by the area/volume of the intersection between the two binary images viewed as random distributions of geometrical objects. Under mild assumptions, it turns out that the appropriately normalized Pearson correlation follows a standard normal distribution under the null hypothesis if the number of image pixels is large. Unlike previous methods, the method handles 2D and 3D images, variable SNRs and any kind of cell shapes. It is able to colocalize large regions with small dots, as it is the case in TIRF-PALM experiments and to detect negative colocalization. The typical processing time is two milliseconds per image pair in 2D and a few seconds in 3D, with no dependence on the number of objects per image. Finally, the method provides maps to geocolocalize molecule interactions in specific image regions.

Collaborators: Jean Salamero and Liu Zengzhen (UMR 144 CNRS-Institut Curie).

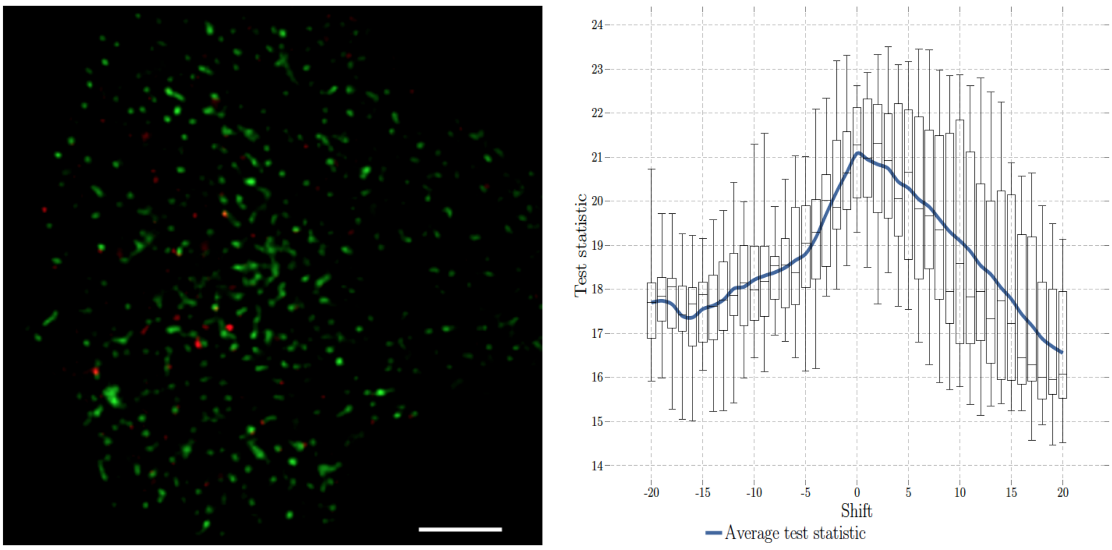


Figure 6. Illustration of statistical spatio-temporal colocalization. a) Average intensity projection of a 3D TIRF acquisition showing m-Cherry Langerin (red channel) and GFP Rab11A proteins (green channel) (courtesy of UMR 144 CNRS-Institut Curie); b) test statistic value with respect to the shift between frames.

7.6. Classification of diffusion dynamics from particle trajectories

Participants: Vincent Briane, Charles Kervrann.

In this study, we are currently interested in describing the dynamics of particles inside live cell. We assume that the motions of particles follow a certain class of random process: the diffusion processes. In 2015, we developed a statistical test to classify the intracellular motions into three groups : free diffusion (*i-e* Brownian motion), subdiffusion and superdiffusion. This method is an alternative to the commonly used Mean Square Displacement (MSD) analysis. This year, we have studied theoretical properties of our procedure. We have shown that it behaves well asymptotically, that is when we observe the particle trajectory for a very long time, for certain parametric models. The models on which we assess our procedure are representative of the three classes aforementioned and extensively used in the literature. Among them we can cite Brownian motion with drift, Ornstein-Uhlenbeck process and fractional Brownian motion. An illustration of the testing procedure is shown in Fig. 7.

We also extend our method to address two different questions. First, we are interested in testing a large number of trajectories. The first version of our test is a single test procedure. It is known that applying multiple times a test without care leads to a high number of false positives. Then, we modify our initial method to overcome this problem. Secondly, in the case in which we observe very long trajectories, it is likely that the particle motion changes over time. Therefore, we are currently adapting our initial procedure to detect change-point along a single trajectory.

Collaborators: Myriam Vimond (ENSAI Rennes),
Jean Salamero (UMR 144 CNRS-Institut Curie).

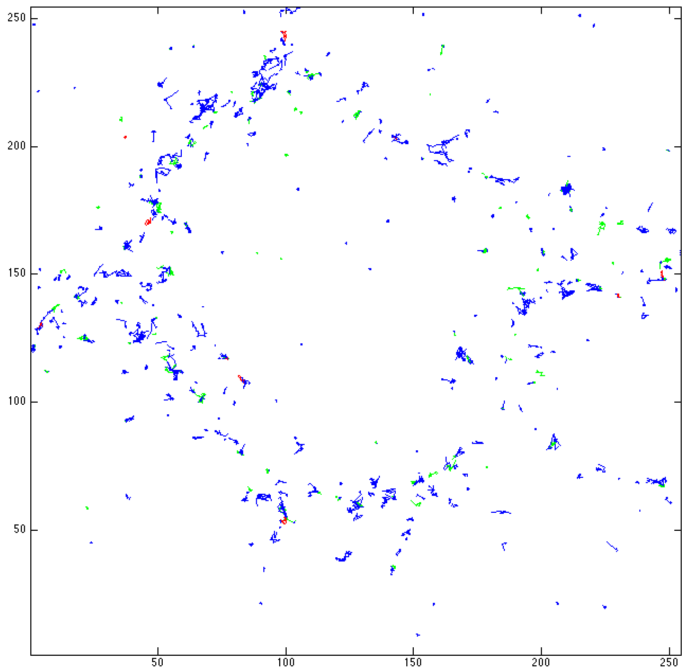


Figure 7. Labelling of the dynamics of trajectories in Single Particle Tracking PALM (Courtesy of Institut Interdisciplinaire de Neurosciences CNRS UMR 5297). The color code is green for subdiffusion, blue for Brownian motion and red for superdiffusion.

7.7. Inference for spatial Gibbs point processes

Participant: Frédéric Lavancier.

Gibbs point processes are popular and widely used models in spatial statistics to describe the repartition of points or geometrical structures in space. They initially arose from statistical physics where they are models for interacting particles. They are now used in as different domains as astronomy, biology, computer science, ecology, forestry, image analysis and materials science.

Assuming a parametric form of the Gibbs interaction, the natural method to estimate the parameters is likelihood inference. Since its first use in the 80's, this method is conjectured to be consistent and efficient. However the theoretical properties of maximum likelihood for Gibbs point processes remain largely unknown. In [17], we partly solved this 30 years old conjecture by proving the consistency of the likelihood procedure for a large class of Gibbs models. As important examples, we deduce the consistency of the maximum likelihood estimator for all parameters of the Strauss model, the hardcore Strauss model, the Lennard-Jones model and the area-interaction model, which are commonly used models in practice.

A practical issue of likelihood estimation yet is that this method depends on an intractable normalizing constant that has to be approximated by simulation. To avoid this problem, other methods of estimation have been introduced, including pseudo-likelihood estimation. The theoretical properties of the pseudo-likelihood method are fairly well known in the case of finite-range Gibbs interactions. However this setting rules out some major Gibbs model as the Lennard-Jones model. In [16], we extend the pseudo-likelihood procedure to infinite range Gibbs interactions and we prove its consistency and its asymptotic normality.

References: [16] [17]

Collaborators: David Dereudre (Laboratoire Paul Painlevé (UMR 8524), University of Lille 1),
Jean-François Coeurjolly (Laboratoire Jean Kutzmann, University of Grenoble).

7.8. Statistical aspects of Determinantal Point Processes

Participant: Frédéric Lavancier.

Determinantal point processes (DPPs) have been introduced in their general form by Macchi (1975) and have been extensively studied from a probabilistic point of view in the 2000's (one of the main reason being their central role in random matrix theory). In a previous work, we have demonstrated that DPPs provide useful models for the description of spatial point pattern datasets where nearby points repel each other.

In [15], we have addressed the question of how repulsive a stationary DPP can be, in order to assess the range of practical situations this promising class of models may model. We determine the most repulsive DPP (in some sense) and we introduce new parametric families of stationary DPPs that can cover a large range of DPPs, from the stationary Poisson process (the case of no interaction) to the most repulsive DPP. Some theoretical aspects of inference for stationary DPPs are tackled in [13] and [14]. In the former study we establish the Brillinger mixing property of stationary DPPs, a first important step toward asymptotic inference. In the latter contribution, we exploit this result to deduce the consistency and asymptotic properties of contrast estimators for stationary DPPs.

References: [13] [15] [14]

Collaborators: Christophe Ange Napoléon Biscio (LMJL, University of Nantes),
Jesper Möller (Department of Mathematical Sciences, Aalborg University, Denmark),
Ege Rubak (Department of Mathematical Sciences, Aalborg University, Denmark).

7.9. Modeling aggregation and regularity in spatial point pattern datasets

Participant: Frédéric Lavancier.

In the spatial point process literature, analysis of spatial point pattern datasets are often classified into three main cases: (i) Regularity (or inhibition or repulsiveness), modelled by Gibbs point processes, hard core processes like Matern hard core models, and determinantal point processes; (ii) Complete spatial randomness, modelled by Poisson point processes; (iii) Aggregation (or clustering), modelled by Poisson cluster processes and Cox processes. For applications the classification (i)-(iii) can be too simplistic, and there is a lack of useful spatial point process models with, loosely speaking, aggregation on the large scale and regularity on the small scale. For instance, we may be interested in such a model for the repartition of the centres of vesicles in a cell, that exhibit some spatial clustering at large scales while having a minimal distance between them.

In [23], we have considered a dependent thinning of a regular point process with the aim of obtaining aggregation on the large scale and regularity on the small scale in the resulting target point process of retained points. Various parametric models for the underlying processes are suggested and the properties of the target point process are studied. Simulation and inference procedures are discussed when a realization of the target point process is observed, depending on whether the thinned points are also observed or not.

Reference: [23]

Collaborator: Jesper Möller (Department of Mathematical Sciences, Aalborg University, Denmark).

7.10. Multi-scale spot segmentation with automatic selection of image scales

Participants: Bertha Mayela Toledo Acosta, Patrick Bouthemy.

Detecting spot-like objects of different sizes in images is required for many applications. A spot detection framework can be divided in three sub-steps : first, image preprocessing to smooth out noise; second, signal enhancement to highlight spots; third, spot detection by thresholding; the two first ones being often merged in a single operator. However, elements of interest do not all correspond to the same image scale, if the collection includes subgroups of different sizes or if perspective effects occur. Then, the need is not merely the selection of the optimal image scale, but of all the meaningful scales. We dealt with the problem of multi-scale spot detection while automatically selecting the meaningful scales. Our primary interest is to detect particles in microscopy images, but our method can be applied to other types of images as well. We defined an original criterion based on the a contrario approach and the LoG scale-space framework to automatically select the meaningful scales. We designed a coarse-to-fine multi-scale spot segmentation scheme involving a locally adaptive thresholding across scales, to come up with the final map of segmented spots. We carried out experimental results on simulated and real images of different types, and we demonstrated that our method outperforms other existing methods.

Reference: paper accepted, ICASSP'2017.
Collaborator: Antoine Basset (CNES, Toulouse).

7.11. Multi-modal registration for correlative light-electron microscopy

Participants: Bertha Mayela Toledo Acosta, Patrick Bouthemy, Charles Kervrann.

We pursue our work on correlative light-electron microscopy (CLEM), which combines the strengths of two different imaging modalities, light microscopy (LM) and electron microscopy (EM), to jointly study intracellular dynamics and ultrastructure of a biological sample. CLEM registration is an important and difficult problem given the significant differences between LM and EM images regarding resolution, field of view, image size and appearance. We designed an automated approach for retracing and registering CLEM images, by implementing a patch-based search using a common Laplacian of Gaussian (LoG) image representation of the LM and EM images. We have redefined the geometry of the patch, opting for a disk-shaped patch. The search (or retracing) step uses histogram-based methods as they are invariant to rotation, and it provides a pre-registration by producing the estimate of the translation component. Usually, there is a large disparity on the orientation of EM and LM images. To handle this problem, we have implemented a mutual information-based method to compute the rotation between the EM and LM patches and to refine the registration. We have also tackled the registration issue in both directions (LM to EM, and EM to LM), and compared our approach to a correlation-based method.

We have tested our approach on a larger set of real CLEM images (provided by Institut Curie) presenting a large diversity in content, image size, and appearance, further validating our method (see Fig. 8). We are currently exploring how our automated CLEM registration method could be exploited to guide EM acquisition within a coarse-to-fine framework.

Reference: [35]
Collaborators: Xavier Heiligenstein (UMR 144 CNRS-Institut Curie),

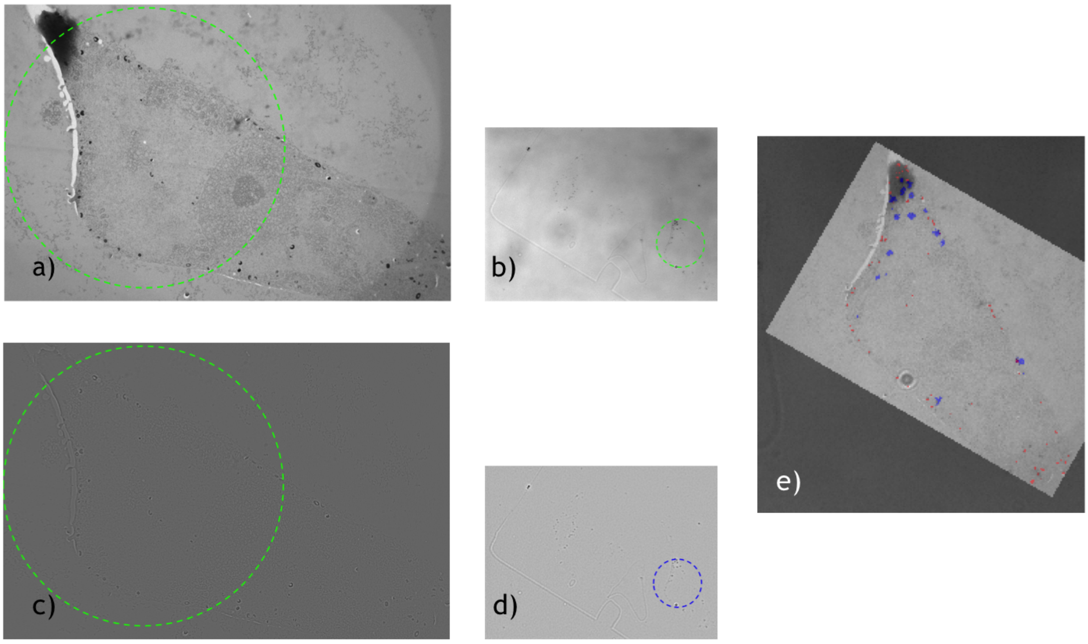


Figure 8. Figure 1. CLEM experiment (images from UMR 144 CNRS-Institut Curie, Xavier Heiligenstein): a) EM image with Region of Interest (ROI) framed in green; b) ground-truth location of the corresponding LM patch framed in green; c) ROI delineated in the Log-EM image, framed in green; d) selected patch (SP) in the LoG-LM image, framed in blue to be comaprd with the green disk in b); f) overlay (after registration).

Grégoire Malandain (Inria, Morpheme EPC, Sophia-Antipolis).

7.12. Denoising and compensation of the missing wedge in cryo electron tomography

Participants: Emmanuel Moebel, Charles Kervrann.

In this study, we address two important issues in cryo electron tomography (CET) images: reduction of noise and restoration of information in the missing wedge (MW). The MW is responsible for several type of imaging artifacts, and arises because of limited angle tomography: it is observable in the Fourier domain and is depicted by a region where Fourier coefficient values are unknown (see Fig. 9). The proposed stochastic method tackles the restoration problem by filling up the MW by iterating following steps : adding noise into the MW (step 1) and applying a denoising algorithm (step 2). The role of the first step is to propose candidates for the missing Fourier coefficients and the second step acts as a regularizer. A constraint is added in the spectral domain by imposing the known Fourier coefficients to be unchanged through iterations.

Several aspects of the method have been studied in order to gain a deeper understanding of this strategy: different kinds of noise as well as several denoising algorithms (BM3D, NL-Bayes, NL-means, Total Variation...) have been evaluated. Furthermore, different kinds of transforms have been tested in order to apply the constraint (Fourier transform, Cosine transform, pseudo-polar Fourier transform). Also, a process has been set up in order to evaluate the performance of the proposed method on experimental data. Thus, convincing results on experimental data have been achieved (see Fig. 9) using the Fourier Shell Correlation (FSC) as an evaluation metric. In order to measure the quality of the recovered MW only, we also compute the FSC over the MW support (“constrained FSC”).

Collaborators: Damien Larivière (Fondation Fourmentin-Guilbert),
Julio Ortiz (Max-Planck Institute, Martinsried, Germany).

7.13. Spatially-variant kernel for optical flow under low signal-to-noise ratios

Participant: Charles Kervrann.

Local and global approaches can be identified as the two main classes of optical flow estimation methods. This year, we have proposed a framework to combine the advantages of these two principles, namely robustness to noise of the local approach and discontinuity preservation of the global approach. The idea is to adapt spatially the local support of the local parametric constraint in the combined local-global model [42]. To this end, we jointly estimate the motion field and the parameters of the spatial support. We apply our approach to the case of Gaussian filtering, and we derive efficient minimization schemes for usual data terms. The estimation of a spatially varying standard deviation map prevents from the smoothing of motion discontinuities, while ensuring robustness to noise. We validated our method for a standard model and demonstrated how a baseline approach with pixel-wise data term can be improved when integrated in our framework. The method has been evaluated on the Middlebury benchmark with ground truth and on real fluorescence microscopy data for which noise is the main limitation for usual optical flow methods.

Collaborator: Denis Fortun (EPFL-BIG, Lausanne, Switzerland)
Noémie Debroux (Laboratory of Mathematics, INSA Rouen, Normandie University)

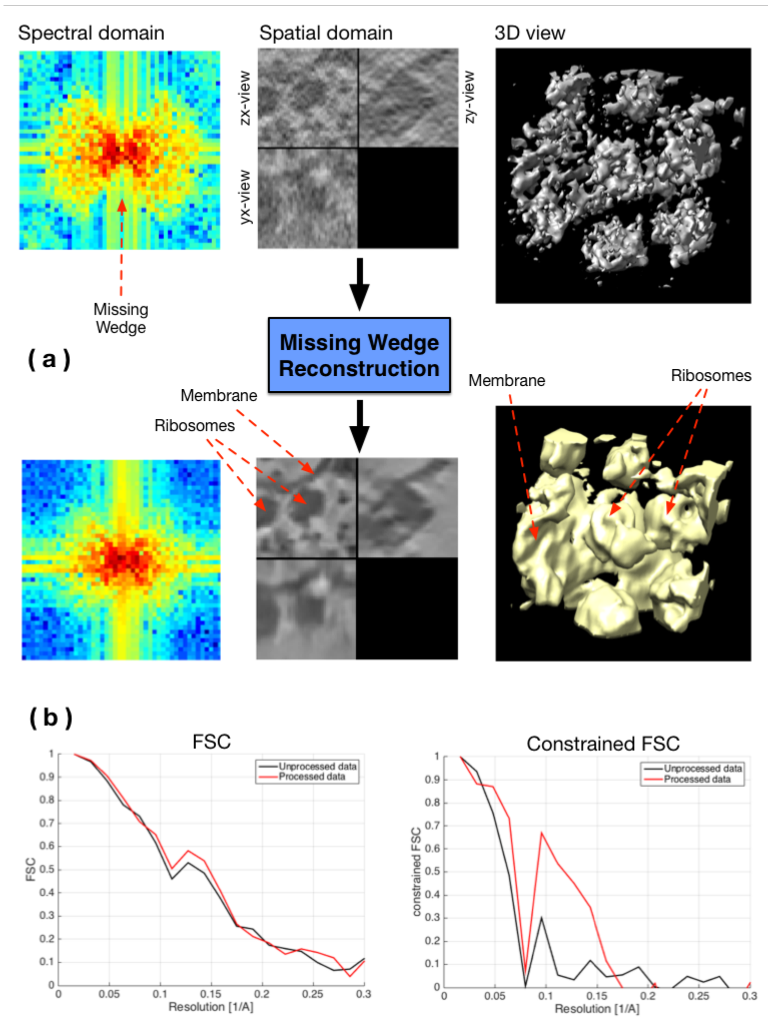


Figure 9. Experimental sub-tomogram containing ribosomes attached to a membrane. (a) Top row: original data in the spectral (left) and spatial (middle) domains and 3D view of the thresholded data (right). Bottom row: denoised data shown as above. (b) FSC and constrained FSC measures of the method input (in black) and output (in red). All measures are wrt the same reference.

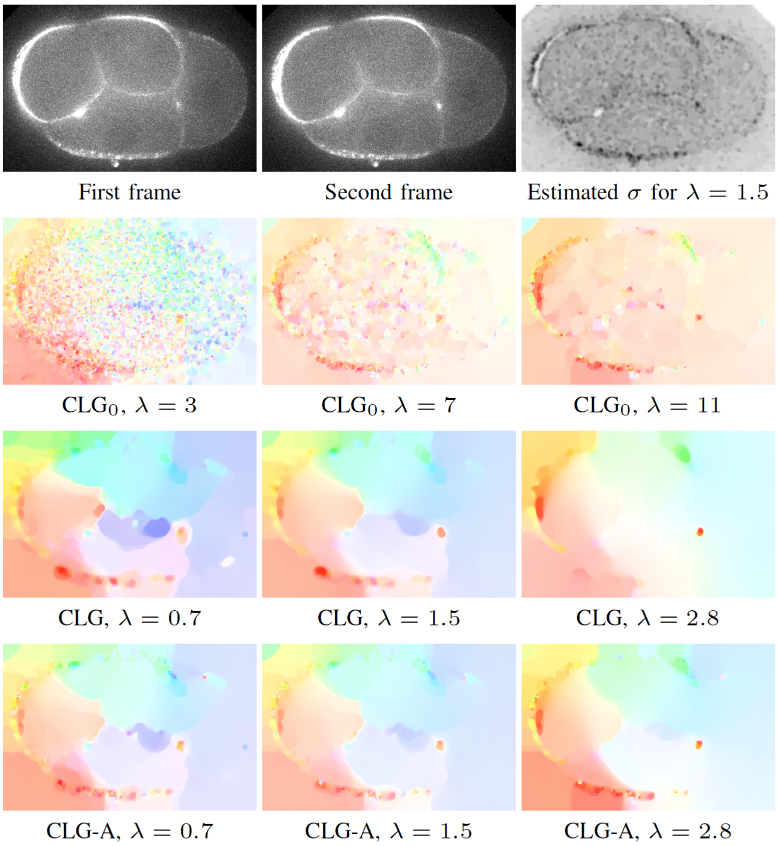


Figure 10. Visual results obtained with variants of CLG ("Combined Local-Global") methods for several values of regularization parameters λ: CLG-A (our adaptive CLG), CLG_0 (pointwise method) and CLG ([42]) on a image pair from a CElegans sequence.

7.14. Frame-based hierarchical motion decomposition and segmentation

Participants: Juan Manuel Perez Rua, Patrick Bouthemy.

A number of applications in video analysis rely on a per-frame motion segmentation of the scene as key preprocessing step. Moreover, different settings in video production require extracting segmentation masks of multiple moving objects and object parts in a hierarchical fashion. In order to tackle this problem, we propose to analyze and exploit the compositional structure of scene motion to provide a segmentation which is not purely driven by local image information. Specifically, we leveraged a hierarchical motion-based partition of the scene to capture a mid-level understanding of the dynamic video content. To recover the decomposition tree, we formulated the problem as a per-pixel label selection interleaved with motion models estimation. The labels represent the set of nodes from the initial proposal tree which are selected to explain globally the input correspondence field. We carried out experimental results showing the strengths of this approach in comparison to current video segmentation approaches. Indeed, they demonstrated the superior ability of our method to capture the main moving objects of the scene in the first layer of the tree, and to segment them in moving parts in deeper layers. As such, we believe our segmentation method is closer to the complex needs of video editing than current hierarchical segmentation approaches.

Reference: [34]
Collaborators: Tomas Crivelli and Patrick Pérez (Technicolor).

7.15. Trajectory-based discovery of motion hierarchies in video sequences

Participants: Juan Manuel Perez Rua, Patrick Bouthemy.

The dynamic content of physical scenes is largely compositional, that is, the movements of the objects and of their parts are hierarchically organized and relate through composition along this hierarchy. This structure also prevails in the apparent 2D motion that a video captures. Visual motion in the scene is roughly organized along a tree, with the dominant motion (typically induced by camera motion) at the root, and motion components adding up along the branches. Accessing this visual motion hierarchy is important to get a better understanding of dynamic scenes and is useful for video manipulation. We proposed to capture it through learned, tree-structured sparse coding of point trajectories. We found that dictionary learning and sparse coding provide appealing tools to disentangle this latent hierarchical structure. More precisely, we introduced a new tree-structured dictionary learning method that allows describing each track with a few basis functions, all but one being inherited from its parent in the structure. The sparse codes thus associated to the tracks capture the desired structure and lend themselves naturally to hierarchical clustering of the collection. We showed through experiments on motion capture data that our model is able to extract moving segments along with their organization. We also obtained competitive results on the task of segmenting objects in real video sequences from trajectories.

Reference: [33]
Collaborators: Tomas Crivelli and Patrick Pérez (Technicolor).

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Contracts with Industry

8.1.1. Contract with Innopsys: Tissue microarrays (TMA) image analysis

Participants: Hoai Nam Nguyen, Charles Kervrann.

Collaborators: Vincent Paveau and Cyril Cauchois (Innopys company).

A three-year contract has been established with Innopsys in 2013 to support Hoai Nam Nguyen's PhD thesis. The objective is to investigate and develop methods and algorithms dedicated to fluorescence images acquired by the scanners and devices designed by the company. In this project, we focus on i/ localization and segmentation fluorescence tissue microarrays (TMA) cores in very large 2D images; ii/ de-arraying of digital images and correction of grid deformation adapted to devices; iii/ correction of scanning artifacts to improve image reconstruction; iv/ deconvolution, denoising and superresolution of fluorescence TMA images corrupted by Poisson noise. The algorithms will be integrated into the plateforms and devices designed by Innopsys.

8.1.2. Contract (CIFRE) with Technicolor: Semantically meaningful motion descriptors for video understanding

Participants: Juan Manuel Perez Rua, Patrick Bouthemy.

Collaborators: Tomas Crivelli and Patrick Pérez (Technicolor).

A three-year contract has been established with Technicolor on January 2015 for a CIFRE grant supporting Juan Manuel Perez Rua's PhD thesis. The purpose is to investigate new methods for extracting meaningful mid-level motion-related descriptors that may help for the semantic discovery of the scene. In 2015, we started with the occlusion detection problem. We have proposed a novel approach where occlusion in the next frame or not is formulated in terms of visual reconstruction. Our approach outperforms state-of-the-art occlusion detection methods on the challenging MPI Sintel dataset. In 2016, we have developed two hierarchical motion segmentation methods involving a compositional motion representation. The first one follows a frame-based labeling approach which amounts to the minimization of a global energy function. The second one is trajectory-based and relies on tree-structured learning and sparse coding.

8.1.3. Contract with OBSYS: microscope set-up control and inverse problems in microscopy

Participants: Giovanni Petrazzuoli, Charles Kervrann.

Collaborators: Charles Gudeudry (OBSYS).

A three-year contract has been established with OBSYS in 2016. The objective is to investigate and develop methods and algorithms dedicated to the control of a microscope set-up and to the analysis of fluorescence images. Fast and robust algorithms will be especially developed to improve image reconstruction of 3D-TIRF microscope images. The algorithms will be integrated into the plateforms and devices designed by OBSYS.

8.2. Bilateral grants with industry

8.2.1. Fourmentin-Guilbert Foundation: Macromolecule detection in cryo electron tomograms

Participants: Emmanuel Moebel, Charles Kervrann.

Collaborator: Damien Larivière (Fourmentin-Guilbert Foundation).

The Fourmentin-Guilbert Foundation strives for building a virtual E. coli bacteria. Information about the position of macromolecules within the cell is necessary to achieve such a 3D molecularly-detailed model. The Fourmentin-Guilbert Foundation supports cutting-edge *in-situ* cryo electron tomography combined with image processing at the Max-Planck Institute of Biochemistry to map the spatial distribution of the ribosomes, and obtain structural information on the complexes they form *in-situ* with cofactors and other ribosomes. The objective of the project is to explore and evaluate novel methods from the field of 3D shape retrieval for identifying, localizing and counting macromolecules (e.g., 70S ribosome) within a tomogram. This project is also supported by "Region Bretagne".

9. Partnerships and Cooperations

9.1. Regional Initiatives

ENSAI-CREST: Statistical methods and models for image registration, Vincent Briane PhD thesis is co-funded by Inria and ENSAI-CREST and co-supervised by Myriam Vimond (ENSAI-CREST).

Région Bretagne: Identification, localization and enumeration of ribosomes within a tomogram by combining state-of-the-art denoising methods and object descriptor-based recognition (CATLAS, see Section 8.2.1).

BioGenOuest: Collaboration with S. Prigent (engineer) in charge of the organization of image processing services for Biogenouest bio-imaging facilities.

IGDR: Collaboration with J. Pecreaux, Y. Le Cunff (co-supervision of A.G. Caranfil's PhD thesis).

9.2. National Initiatives

9.2.1. France-BioImaging project

Participants: Charles Kervrann, Patrick Bouthemy, Thierry Pécot, Emmanuel Moebel, Ancageorgiana Caranfil.

The goal of the project is to build a distributed coordinated French infrastructure for photonic and electronic cellular bioimaging dedicated to innovation, training and technology transfer. High-computing capacities are needed to exhaustively analyse image flows. We address the following problems: i/ exhaustive analysis of bioimaging data sets; ii/ deciphering of key steps of biological mechanisms at organ, tissular, cellular and molecular levels through the systematic use of time-lapse 3D microscopy and image processing methods; iii/ storage and indexing of extracted and associated data and metadata through an intelligent data management system. SERPICO is co-head of the IPDM (Image Processing and Data Management) node of the FBI network composed of 6 nodes.

Funding: Investissement d'Avenir - Infrastructures Nationales en Biologie et Santé ANR (2011-2016).

Partners: CNRS, Institut Jacques Monod, Institut Pasteur, Institut Curie, ENS Ulm, Ecole Polytechnique, INRA, INSERM.

9.2.2. ANR DALLISH project (2016-2020): Data Assimilation and Lattice LIght SHeet imaging for endocytosis/exocytosis pathway modeling in the whole cell

Participants: Charles Kervrann, Patrick Bouthemy, Vincent Briane, Ancageorgiana Caranfil.

The Lattice Light Sheet Microscopy (LLS-M) represents at present the novel generation of 3D fluorescence microscopes dedicated to single cell analysis, generating extraordinarily high resolved and sharp, but huge 3D images and videos: one single live cell experiment in one single condition, imaging two molecular markers of the endocytosis pathway and using cutting-edge LLS-M can result into up to one Terabyte of data, at the spatial resolution of 100-200 nanometers in 3D. In such a situation, it is found the usual conventional image reconstruction algorithms and image analysis methods developed for 3D fluorescence microscopy are likely to fail to process a deluge of voxels generated by LLS-M instruments. The goal of the project is then to develop new paradigms and computational strategies for image reconstruction and 3D molecule tracking/motion estimation. Furthermore, establishing correspondences between the image-based measurements and features (e.g., motion vectors, trajectories), stochastic motion models and the underlying biological and biophysical information remains a challenging task.

The impact of the project will be three-fold. First, our new image processing paradigms and improved algorithms (allowing faster, more resolved and more accurate results) will have direct benefits in modern bioimaging. Second, the methods and algorithms will apply to decipher molecular mechanisms of protein transports, here focused on endocytosis/exocytosis. Finally, in a larger perspective, the quantitative description of protein transport will be a prerequisite for understanding the functioning of a cell in normal and pathological situations, as default in protein transport appeared over the years, as a major contributory factor to a number of diseases, including cancer, viral infection and neurodegenerative diseases.

Funding: ANR - Agence Nationale de la Recherche

Partners: Inria (SERPICO, BEAGLE, Fluminance), INRA MaIAGE Unit Jouy-en-Josas, Institut Curie (UMR CNRS 144 & U1143 Inserm UMR 3666) Paris

9.3. European Initiatives

9.3.1. Major European Organizations with which the Team have followed Collaborations

ESFRI Euro-BioImaging initiative: SERPICO participates in the ESFRI Euro-BioImaging project, one of the four new biomedical science projects in the roadmap of the European Strategic Forum on Research Infrastructures (ESFRI). The mission of Euro-BioImaging is to provide access, service and training to state-of-the-art imaging technologies and foster the cooperation and networking at the national and European level including multidisciplinary scientists, industry, regional, national and European authorities. SERPICO also participates to the French counterpart, the so-called "France-BioImaging" (FBI) network which gathers several outstanding cellular imaging centers (microscopy, spectroscopy, probe engineering and signal processing) as described in Section 9.2.1.

9.4. International Initiatives

9.4.1. Informal International Partners

Collaboration with Max-Planck Institute, Martinsried (Germany), Dr. Julio Ortiz: Detection and segmentation of macromolecules in cryo electron tomography (project in progress with Emmanuel Moebel and Charles Kervrann).

Collaboration with Aalborg University (Denmark), Prof. Rasmus Waagepetersen : Estimating equations for inhomogeneous determinantal point processes (project in progress with Frédéric Lavancier).

Collaboration with EPFL (M. Unser's Team, Switzerland). D. Fortun: optical flow computing (project in progress with Charles Kervrann).

9.4.2. CytoDI Inria Associated-Team

Title: Quantitative Imaging of Cytoskeleton Dynamics in 3D

International Partner:

University of Texas, SouthWestern Medical Center, Dallas (United States) - Gaudenz Danuser

Start year: 2016

See also: http://serpico.rennes.inria.fr/doku.php?id=research:cytodi

The main scientific goal of the Associated-Team is the spatiotemporal characterization and comparison of cytoskeleton networks involved in cell migration and observed through live cell imaging in three dimensions (3D). Those networks include the cytoskeleton, i.e., microtubules (MT), intermediate filaments (IF), dynamically resolvable by Bessel Beam Light Sheet fluorescent microscopy. The goal will be achieved through the design of local and global descriptors of the spatial conformation and deformation of the cytoskeleton. Subsequently, general metrics to compare and classify the MT and IF networks will be investigated. This study will be carried out on oncogenically transformed lung cancer epithelial cells.

The first meeting took place in Dallas in May 2016 as originally scheduled, to discuss and update current research direction and discuss scientific progress.

9.5. International Research Visitors

9.5.1. Visits to International Teams

Visit of 3 months of Juan Manuel Perez Rua in the Philip Torr's team (University of Oxford, UK).
Visit of 1 one week of Vincent Briane to the ESGI (European Study Group in Industry) in Dublin (Ireland, July 2016).

Visit of 1 one week of Vincent Briane to the University of Limerick (K. Burke's team) (Ireland, November 2016).

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events Organisation

10.1.1.1. Member of the Organizing Committees

Frédéric Lavancier and Charles Kervrann were head of the organizing committee of the international workshop "Spatial Statistics and Image Analysis in Biology" (SSIAB'16), Inria Rennes, May 2016.

Frédéric Lavancier was head of the workshop "Spatio-Temporal Models and Statistics", IRMAR University of Rennes 1, LMJL University of Nantes, ENSAI, University of Rennes 2, INRA Rennes, Inria Rennes, April 2016.

Patrick Bouthemy and Thierry Pécot were respectively main organizer and co-organizer of the BioImage Computing workshop in conjunction with ECCV'2016, Amsterdam, October 2016.

Patrick Bouthemy and Charles Kervrann were respectively main organizer and co-organizer of the mini-symposium "Image analysis advances in dynamic microscopy and live cell imaging" in SIAM Conference on Imaging Sciences, Albuquerque, New-Mexico, USA, May 2016.

Charles Kervrann was member of the organization committee of the Microscopy school MiFoBio'2016, Seignosse, October 2016.

10.1.2. Scientific Events Selection

10.1.2.1. Member of the Conference Program Committees

Charles Kervrann: member of the scientific committee of RFIA'2016 (Clermont-Ferrand), Associated Editor for the conference ISBI'2017, member of the scientific committee of the "Image the Cell 2017" conference (Rennes), member of the scientific committee "Journées d'Imagerie Optique Non-Conventionnelle" (JIONC'2016 and JIONC'2017).

Patrick Bouthemy: Associate Editor for the conference ISBI'2016 and ISBI'2017, Area Chair of ICIP'2016, member of the program committee of RFIA'2016.

Thierry Pécot: member of the program committee of BIOIMAGING'2016.

10.1.2.2. Reviewer

Charles Kervrann: reviewer for ICIP'2016, ICASSP'2016, ISBI'2016, ICASSP'2017, ICIP'2017.

Patrick Bouthemy: reviewer for ICIP'2016, ISBI'2016.

Thierry Pécot: reviewer for ISBI'2016, BIOIMAGING'2016.

10.1.3. Journal

10.1.3.1. Member of the Editorial Boards

Charles Kervrann is Guest Editor of the special issue entitled "Advanced Signal Processing in Microscopy and Cell Imaging" of the IEEE Selected Topics in Signal Processing Journal, February 2016.

Charles Kervrann is Associate Editor of the IEEE Signal Processing Letters journal (since January 2015).

Patrick Bouthemy is co-editor in chief of the open access journal Frontiers in ICT, specialty Computer Image Analysis.

10.1.3.2. Reviewer - Reviewing Activities

Charles Kervrann: BMC Bioinformatics, IEEE Transactions on Computational Imaging, SIAM J. on Imaging Sciences, Methods, expert for the project evaluation in the framework of FONDECYT Science Council (Chile).

Patrick Bouthemy: IEEE Transactions on Image Processing, Mathematical Problems in Engineering, IEEE Signal Processing Letters, IEEE Robotics and Automation Letters, Applied Soft Computing Journal, Computational Intelligence and Neuroscience.

Frédéric Lavancier: Spatial Statistics, R Journal

Thierry Pécot: Bioinformatics, IEEE Transactions on Medical Imaging.

10.1.4. Invited Talks

Charles Kervrann:

Invited talk at the FBI seminar (Paris centre), Computational methods for diffusion, motion and molecular interaction estimation, Pont l'Eveque, February 2016.

Invited talk at the Forum Mathématique Diderot, Computational analysis of intracellular membrane dynamics: from live cell images to biophysical model, Paris, March 2016.

Seminar UTSW, Danuser's lab, Computational analysis of intracellular membrane dynamics: from live cell images to biophysical model, Dallas, TX, USA, May 2016.

Invited talk at SIAM Conference on Imaging Sciences, PEWA: Patch-based Exponentially Weighted Aggregation for image denoising, Albuquerque, New-Mexico, USA, May 2016.

Patrick Bouthemy

Invited talk at SIAM Conference on Imaging Sciences, A scale-adaptive method for retracing and registering in correlative light-electron microscopy, Albuquerque, New-Mexico, USA, May 2016.

Vincent Briane:

Invited talk at SSIAB'2016 workshop, An adaptive statistical test to detect non Brownian diffusion from particle trajectories, Inria Rennes, May 2016.

Thierry Pécot:

Invited talk at SSIAB'2016 workshop, A non-parametric procedure for co-localization studies in fluorescence microscopy, Inria Rennes, May 2016.

Talk and practical course at Microscopy school MiFoBio'2016, mage processing methods for the temporal analysis of moving particles, Seignosse, October 2016.

Seminar IGDR, QuantEv: Quantifying the spatial distribution of intracellular events, Rennes, November 2016.

Frédéric Lavancier:

Seminar of Statistics and Probability in Lille 1, Determinantal point process models and statistical inference, Lille, March 2016.

Seminar in Statistics of University Toulouse 1, A general procedure to combine estimators, Toulouse, June 2016.

Invited talk at "Journées MAS", Determinantal point process models and statistical inference, Grenoble, August 2016.

10.1.5. Scientific Expertise

Charles Kervrann:

Member of the IEEE BISP "Biomedical Image and Signal Processing" committee.

Member of executive board of the GdR MIV (2588 - Microscopie Fonctionnelle du Vivant) CNRS.

Member of the scientific committee of the Interdisciplinary MiFoBio School CNRS (http://www.mifobio.fr).

Patrick Bouthemy:

Member of the board of AFRIF (Association Française pour la Reconnaissance et l'Interprétation des Formes).

Member of the Research Committee of Telecom Bretagne.

Frédéric Lavancier:

Elected member of CNU section 26.

Expert for the project evaluation in the framework of FNRS (Fonds de la Recherche Scientifique), Belgium.

10.1.6. Research Administration

Charles Kervrann:

Member of the executive board of the project committee of the Inria Rennes - Bretagne Atlantique centre.

Member of the Scientific Council of the INRA Rennes Research Centre.

Co-head of the “BioImage Informatics” node (ANR France-BioImaging project, National Infrastructure en Biologie et Santé).

Patrick Bouthemy:

Head of Excellence Lab CominLabs since April 2014.

Deputy member of the board of directors and member of the Selection and Validation Committee of the Images & Réseaux competitivity cluster.

Deputy member of the board of directors of IRT (Technological Research Institute) b<>com.

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

Charles Kervrann:

Engineer Degree: Genomics and Informatics, 4.5 hours, Ecole Nationale Supérieure des Mines de Paris.

Master: From Bioimage Processing to BioImage Informatics, 5 hours, coordinator of the module (30 hours), Master 2 Research IRIV, Telecom-Physique Strasbourg and University of Strasbourg.

Master: Geometric Modeling for Shapes and Images, 6 hours, Master 2 Research SISEA, University of Rennes 1.

Engineer Degree and Master 2 Statistics and Mathematics: Statistical Models and Image Analysis, 37 hours + 15 hours (TP, Emmanuel Moebel), 3rd year, Ecole Nationale de la Statistique et de l’Analyse de l’Information (ENSAI), Rennes.

Patrick Bouthemy:

Master: Analysis of Image Sequences, 18 hours, Master 2 Research SISEA, ISTIC & University of Rennes 1.

Master: Video Indexing, 9 hours, Master 2 Research Computer Science, ISTIC & University of Rennes 1.

Engineer Degree and Master 2 Research IRIV: Motion Analysis, 12 hours, Telecom-Physique Strasbourg & University of Strasbourg.

Frédéric Lavancier:

Master: Linear Models, 36 hours, Master 2 Mathematics & Engineering, option Statistics, University of Nantes.

Master: Time Series, 36 hours, Master 2 Mathematics & Engineering, option Statistics, University of Nantes.

10.2.2. Supervision

PhD in progress: Arnaud Poinas, Inference for inhomogeneous determinantal point processes, started in September 2016, supervised by Bernard Delyon and Frédéric Lavancier

PhD in progress: Hoai Nam Nguyen, Methods and algorithms for tissue microarrays image analysis, started in October 2013, supervised by Charles Kervrann and Vincent Paveau (Innopsys).

PhD in progress: Vincent Briane, Statistical methods and models for motion analysis in microscopy, started in October 2014, supervised by Charles Kervrann and Myriam Vimond (ENSAI-CREST).

PhD in progress: Bertha Mayela Toledo Acosta, Methods and algorithms for 3D image registration and correlative microscopy, started in October 2014, supervised by Patrick Bouthemy and Charles Kervrann.

PhD in progress: Emmanuel Moebel, New strategies for the nonambiguous identification and enumeration of macromolecules in cryo electron tomograms, started in November 2015, supervised by Charles Kervrann.

PhD in progress: Juan Manuel Perez Rua, Semantically meaningful motion descriptors for video understanding, started in January 2015, supervised by Patrick Bouthemy in collaboration with Tomas Crivelli and Patrick Pérez (Technicolor).

PhD in progress: Ancageorgiana Caranfil, Data assimilation methods for cell division mechanisms and molecule trafficking analysis, started in December 2016, supervised by Charles Kervrann and Yann Le Cunff.

PhD in progress: Sandeep Manandhar, Optical flow methods for 3D fluoresence imaging, started in October 2016, supervised by Patrick Bouthemy and Charles Kervrann.

10.2.3. Juries

Referee of PhD thesis: G. Michelin (University of Côte d'Azur, supervised by G. Malandain) [P. Bouthemy], L. Azzari (Tampere University of Technology, Finland, supervised by A. Foi) [C. Kervrann], H. Robjani (University of Strasbourg, supervised by C. Ronse) [C. Kervrann].

11. Bibliography

Major publications by the team in recent years

[1] A. Basset, J. Boulanger, J. Salamero, P. Bouthemy, C. Kervrann.*Adaptive spot detection with optimal scale selection in fluorescence microscopy images*, in "IEEE Transactions on Image Processing", November 2015, vol. 24, n^{o} 11, 16 [*DOI :* 10.1109/TIP.2015.2450996], https://hal.inria.fr/hal-01248290.

[2] J. Boulanger, A. Gidon, C. Kervrann, J. Salamero.*A patch-based method for repetitive and transient event detection in fluorescence imaging*, in "PLoS ONE", Oct 2010, vol. 5, n^{o} 10 [*DOI :* 10.1371/journal.pone.0013190].

[3] J. Boulanger, C. Kervrann, P. Bouthemy.*Space-time adaptation for patch based image sequence restoration*, in "IEEE Transactions on Pattern Analysis and Machine Intelligence", 2007, vol. 29, n^{o} 6, p. 1096–1102.

[4] J. Boulanger, C. Kervrann, P. Bouthemy.*A simulation and estimation framework for intracellular dynamics and trafficking in video-microscopy and fluorescence imagery*, in "Medical Image Analysis", 2009, vol. 13, p. 132–142.

[5] J. Boulanger, C. Kervrann, P. Bouthemy, P. Elbau, J.-B. Sibarita, J. Salamero.*Patch-based nonlocal functional for denoising fluorescence microscopy image sequences*, in "IEEE Transactions on Medical Imaging", Feb 2010, vol. 29, n^{o} 2, p. 442-453 [*DOI :* 10.1109/TMI.2009.2033991].

[6] P.M. CARLTON, J. BOULANGER, C. KERVRANN, J.-B. SIBARITA, J. SALAMERO, S. GORDON-MESSER, J. HABER, S. HAASE, L. SHAO, L. WINOTO, A. MATSUDA, P. KNER, S. USAWA, Y. STRUKOV, M. GUSTAFSSON, Z. KAM, D. AGARD, J.W. SEDAT.*Fast live simultaneous multiwavelength four-dimensional optical microscopy*, in "Proc Natl Acad Sci USA", Sep 2010, vol. 107, n^{o} 37, p. 16016-16022 [*DOI :* 10.1073/PNAS.1004037107].

[7] T. CRIVELLI, B. CERNUSCHI-FRIAS, P. BOUTHEMY, J.-F. YAO.*Motion Textures: Modeling, Classification, and Segmentation Using Mixed-State*, in "SIAM Journal on Imaging Sciences", December 2013, vol. 6, n^{o} 4, p. 2484-2520 [*DOI :* 10.1137/120872048], https://hal.inria.fr/hal-00931667.

[8] A. GIDON, S. BARDIN, B. CINQUIN, J. BOULANGER, F. WAHARTE, L. HÉLIOT, H. DE LA SALLE, D. HANAU, C. KERVRANN, B. GOUD, J. SALAMERO.*A Rab11A/myosin Vb/Rab11-FIP2 complex frames two late recycling steps of langerin from the ERC to the plasma membrane*, in "Traffic", April 2012, vol. 13, n^{o} 6, p. 815-833 [*DOI :* 10.1111/J.1600-0854.2012.01354.X], https://hal.inria.fr/hal-00782005.

[9] C. KERVRANN, J. BOULANGER.*Optimal spatial adaptation for patch-based image denoising*, in "IEEE Transactions on Image Processing", 2006, vol. 15, n^{o} 10, p. 2866–2878.

[10] C. KERVRANN, J. BOULANGER.*Local adaptivity to variable smoothness for exemplar-based image denoising and representation*, in "International Journal of Computer Vision", August 2008, vol. 79, n^{o} 1, p. 45–69.

[11] C. KERVRANN.*PEWA: Patch-based Exponentially Weighted Aggregation for image denoising*, in "NIPS - Neural Information Processing Systems", Montreal, Canada, Neural Information Processing Systems Foundation, December 2014, https://hal.inria.fr/hal-01103358.

[12] T. PÉCOT, P. BOUTHEMY, J. BOULANGER, A. CHESSEL, S. BARDIN, J. SALAMERO, C. KERVRANN.*Background Fluorescence Estimation and Vesicle Segmentation in Live Cell Imaging with Conditional Random Fields*, in "IEEE Transactions on Image Processing", February 2015, vol. 24, n^{o} 2, 14 [*DOI :* 10.1109/TIP.2014.2380178], https://hal.inria.fr/hal-01103126.

Publications of the year

Articles in International Peer-Reviewed Journal

[13] C. A. N. BISCIO, F. LAVANCIER.*Brillinger mixing of determinantal point processes and statistical applications*, in "Electronic journal of statistics ", March 2016, vol. 10, n^{o} 1, p. 582-607, https://hal.archives-ouvertes.fr/hal-01179831.

[14] C. A. N. BISCIO, F. LAVANCIER.*Contrast estimation for parametric stationary determinantal point processes*, in "Scandinavian Journal of Statistics", 2016, https://hal.archives-ouvertes.fr/hal-01215582.

[15] C. A. N. BISCIO, F. LAVANCIER.*Quantifying repulsiveness of determinantal point processes*, in "Bernoulli", May 2016, vol. 22, n^{o} 4, p. 2001-2028, https://hal.archives-ouvertes.fr/hal-01003155.

[16] J.-F. COEURJOLLY, F. LAVANCIER.*Parametric estimation of pairwise Gibbs point processes with infinite range interaction*, in "Bernoulli", January 2016, https://hal.archives-ouvertes.fr/hal-01092225.

[17] D. DEREUDRE, F. LAVANCIER.*Consistency of likelihood estimation for Gibbs point processes*, in "The Annals of Statistics", 2016, https://hal.archives-ouvertes.fr/hal-01144877.

[18] D. FORTUN, P. BOUTHEMY, C. KERVRANN.*A Variational Aggregation Framework for Patch-Based Optical Flow Estimation*, in "Journal of Mathematical Imaging and Vision", 2016, vol. 56, p. 280 - 299 [*DOI :* 10.1007/S10851-016-0664-6], https://hal.archives-ouvertes.fr/hal-01408771.

[19] D. FORTUN, P. BOUTHEMY, C. KERVRANN.*Aggregation of local parametric candidates with exemplar-based occlusion handling for optical flow*, in "Computer Vision and Image Understanding", January 2016, 17 [*DOI :* 10.1016/J.CVIU.2015.11.020], https://hal.inria.fr/hal-01001758.

[20] M. JAIN, H. JÉGOU, P. BOUTHEMY.*Improved Motion Description for Action Classification* , in "Frontiers in information and communication technologies", January 2016 [*DOI :* 10.3389/FICT.2015.00028], https://hal.inria.fr/hal-01401833.

[21] C. KERVRANN, S. T. ACTON, J.-C. OLIVO-MARIN, C. O. S. SORZANO, M. UNSER.*Introduction to the issue on Advanced Signal Processing in Microscopy and Cell Imaging*, in "IEEE Journal of Selected Topics in Signal Processing", 2016, vol. 10, n^{o} 1, p. 3-5 [*DOI :* 10.1109/JSTSP.2015.2511299], https://hal.archives-ouvertes.fr/hal-01251394.

[22] C. KERVRANN, C. O. S. SORZANO, S. T. ACTON, J.-C. OLIVO-MARIN, M. UNSER.*A Guided Tour of Selected Image Processing and Analysis Methods for Fluorescence and Electron Microscopy*, in "IEEE Selected Topics in Signal Processing", February 2016, vol. 10, n^{o} 2, p. 6-30, https://hal.inria.fr/hal-01246375.

[23] F. LAVANCIER, J. MØLLER.*Modelling aggregation on the large scale and regularity on the small scale in spatial point pattern datasets*, in "Scandinavian Journal of Statistics", 2016, vol. 43, n^{o} 2, p. 587-609, https://hal.archives-ouvertes.fr/hal-01155646.

[24] F. LAVANCIER, P. ROCHET.*A general procedure to combine estimators*, in "Computational Statistics and Data Analysis", 2016, vol. 94, p. 175-192, https://hal.archives-ouvertes.fr/hal-00936024.

[25] T. PÉCOT, C. KERVRANN, J. SALAMERO, J. BOULANGER.*Counting-based particle flux estimation for traffic analysis in live cell imaging*, in "IEEE Journal of Selected Topics in Signal Processing", February 2016, vol. 10, n^{o} 1, p. 203-216, https://hal.inria.fr/hal-01417147.

Invited Conferences

[26] V. BRIANE, M. VIMOND, C. KERVRANN.*An adaptive statistical test to detect non Brownian diffusion from particle trajectories*, in "Workshop SSIAB - Spatial Statistics and Image Analysis in Biology", Rennes, France, May 2016, https://hal.inria.fr/hal-01416812.

[27] C. KERVRANN.*PEWA: Patch-Based Exponentially Weighted Aggregation for Image Denoising*, in "SIAM Imaging Science", Albuquerque, United States, SIAM, May 2016, https://hal.inria.fr/hal-01416847.

[28] T. PÉCOT, F. LAVANCIER, C. KERVRANN.*A non-parametric procedure for co-localization studies in fluorescence microscopy*, in "Workshop SSIAB - Spatial Statistics and Image Analysis in Biology", Rennes, France, May 2016, https://hal.inria.fr/hal-01416844.

[29] B. M. TOLEDO ACOSTA, P. BOUTHEMY, C. KERVRANN.*A scale-adaptive method for retracing and registering in correlative light-electron microscopy*, in "SIAM Imaging Science", Albuquerque, United States, SIAM, May 2016, https://hal.inria.fr/hal-01416850.

International Conferences with Proceedings

[30] P. BOUTHEMY, B. M. TOLEDO ACOSTA, B. DELYON.*Robust selection of parametric motion models in image sequences*, in "2016 IEEE International Conference on Image Processing (ICIP)", Phoenix, United States, October 2016, p. 3743 - 3747 [*DOI :* 10.1109/ICIP.2016.7533059], https://hal.inria.fr/hal-01400895.

[31] V. BRIANE, M. VIMOND, C. KERVRANN.*An adaptive statistical test to detect non Brownian diffusion from particle trajectories*, in "ISBI 2016 - IEEE Symposium on Biomedical Imaging", Prague, Czech Republic, IEEE, April 2016, p. 1-4, https://hal.inria.fr/hal-01416855.

[32] J.-M. PÉREZ-RÚA, T. CRIVELLI, P. BOUTHEMY, P. PÉREZ.*Determining occlusions from space and time image reconstructions*, in "2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)", Las Vegas, United States, June 2016, Peer-reviewed paper accepted for presentation at the IEEE Conference on Computer Vision and Pattern Recognition 2016, in Las Vegas, USA, https://hal.archives-ouvertes.fr/hal-01307703.

[33] J.-M. PÉREZ-RÚA, T. CRIVELLI, P. PÉREZ, P. BOUTHEMY.*Discovering motion hierarchies via tree-structured coding of trajectories*, in "Proceedings of the British Machine Vision Conference 2016", York, United Kingdom, BMVA, September 2016, https://hal.archives-ouvertes.fr/hal-01358454.

[34] J.-M. PÉREZ-RÚA, T. CRIVELLI, P. PÉREZ, P. BOUTHEMY.*Hierarchical Motion Decomposition for Dynamic Scene Parsing* , in "2016 IEEE International Conference on Image Processing (ICIP)", Phoenix, United States, IEEE, September 2016, Peer-reviewed paper accepted for presentation at the IEEE International Conference on Image Processing 2016, https://hal.archives-ouvertes.fr/hal-01314095.

[35] B. M. TOLEDO ACOSTA, P. BOUTHEMY, C. KERVRANN.*A Common Image Representation And A Patch-Based Search For Correlative Light-Electron-Microscopy (Clem) Registration*, in "2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)", Prague, Czech Republic, April 2016, p. 257- 260 [*DOI :* 10.1109/ISBI.2016.7493258], https://hal.inria.fr/hal-01400854.

National Conferences with Proceeding

[36] A. BASSET, P. BOUTHEMY, J. BOULANGER, F. WAHARTE, J. SALAMERO, C. KERVRANN.*Modélisation et estimation de la diffusion de protéines à l'exocytose dans des séquences d'images de microscopie TIRF*, in "Reconnaissance des Formes et l'Intelligence Artificielle (RFIA'16)", Clermont-Ferrand, France, Association Française de Reconnaissance et d'Interprétation des Formes (AFRIF), Association Française d'Intelligence Artificielle (AFIA), June 2016, p. 1-8, https://hal.inria.fr/hal-01416868.

[37] V. BRIANE, C. KERVRANN, M. VIMOND.*Test statistique pour détecter les diffusions non browniennes*, in "48e Journées de Statistique", Montpellier, France, SFdS, May 2016, p. 1-6, https://hal.inria.fr/hal-01416862.

Conferences without Proceedings

[38] E. MOEBEL, C. KERVRANN.*Denoising and compensation of the missing wedge in cryo electron tomography*, in "16th European Microscopy Congress", Lyon, France, European Microscopy Society, August 2016, https://hal.inria.fr/hal-01416801.

Other Publications

[39] F. LAVANCIER, P. ROCHET.*A tutorial on estimator averaging in Spatial Statistics models*, July 2016, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01341125.

References in notes

[40] D. AXELROD.*Total Internal Reflection Fluorescent Microscopy in cell biology*, in "Traffic", 2004, vol. 2, p. 4658–4668.

[41] E. BETZIG, G. PATTERSON, R. SOUGRAT, O. LINDWASSER, S. OLENYCH, J. BONIFACINO, M. DAVIDSON, J. LIPPINCOTT-SCHWARTZ, H. HESS.*Imaging intracelluar fluorescent proteins at nanometer resolution*, in "Science", 2006, vol. 313, p. 1642–1645.

[42] A. BRUHN, J. WEICKERT, C. SCHNÖRR.*Lucas/Kanade meets Horn/jaqaSchunck: combining local and global optic flow methods*, in "International Journal of Computer Vision", 2005, vol. 61, n^{o} 3, p. 211–231.

[43] A. BUADES, B. COLL, J.-M. MOREL.*A review of image denoising algorithms, with a new one*, in "SIAM J. Multiscale Modeling & Simulation", 2005, vol. 4, n^{o} 2, p. 490–530.

[44] A. CHESSEL, B. CINQUIN, S. BARDIN, J. BOULANGER, J. SALAMERO, C. KERVRANN.*A detection-based framework for the analysis of recycling in TIRF microscopy*, in "IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI 2010)", Pays-Bas Rotterdam, IEEE Signal Processing Society, Apr 2010, p. 1281-1284 [*DOI :* 10.1109/ISBI.2010.5490230].

[45] A. CHESSEL, B. CINQUIN, S. BARDIN, J. SALAMERO, C. KERVRANN.*Computational geometry-based scale-space and modal image decomposition: application to light video-microscopy imaging*, in "Conf. on Scale Space and Variational Methods (SSVM'09)", Voss, Norway, June 2009, p. 770–781.

[46] T. CORPETTI, E. MÉMIN, P. PÉREZ.*Dense estimation fluid flows*, in "IEEE Trans. Pattern Analysis and Machine Intelligence", 2002, vol. 24, n^{o} 3, p. 365–380.

[47] T. CRIVELLI, P. BOUTHEMY, B. CERNUSCHI-FRIAS, J.-F. YAO.*Simultaneous motion detection and background reconstruction with a conditional mixed-state Markov random field*, in "International Journal of Computer Vision", 2011, vol. 94, n^{o} 3, p. 295–316.

[48] A. GENOVESIO, T. LIEDL, V. EMILIANI, W. PARAK, M. COPPEY-MOISAN, J.-C. OLIVO-MARIN.*Multiple particle tracking in 3D+t microscopy: method and application to the tracking of endocytosed quantum dots*, in "IEEE Trans. Image Processing", 2006, vol. 15, n^{o} 5, p. 1062–1070.

[49] M. GUSTAFSSON, L. SHAO, P.M. CARLTON, R. WANG, I. GOLUBOVSKAYA, W. CANDE, D. AGARD, J.W. SEDAT.*3D resolution doubling in wide-filed fluorescence microscopy by structured illumination*, in "Biophysical J.", 2008, vol. 94, p. 4957–4970.

[50] H. HAUSSECKER, D. FLEET.*Computing optical flow with physical models of brightness variation*, in "IEEE Trans. Pattern Analysis and Machine Intelligence", 2001, vol. 23, n^{o} 6, p. 661–673.

[51] T. KLAR, S. JAKOBS, M. DYBA, A. EGNER, S. HELL.*Fluorescence microscopy with diffraction resolution barrier broken by stimulated emission*, in "Proc. Natl. Acad. Sci. USA", 2000, vol. 97, p. 8206–8210.

[52] C. LAVELLE, H. BERRY, G. BESLON, F. GINELLI, J. GIAVITTO, Z. KAPOULA, A. LE BIVIC, N. PEYRIERAS, O. RADULESCU, A. SIX, V. THOMAS-VASLIN, P. BOURGINE.*From molecules to organisms: towards multiscale integrated model of biological systems*, in "Theoretical Biology Insights", 2008, vol. 1, p. 13–22.

[53] F. NÉDÉLEC.*Computer simulations reveal motor properties generating stable antiparallel microtubule interactions*, in "The Journal of Cell Biology", 2002, vol. 158, n^o 6, p. 1005–1015.

[54] S. OZERE, P. BOUTHEMY, F. SPINDLER, P. PAUL-GILLOTEAUX, C. KERVRANN.*Robust parametric stabilization of moving cells with intensity correction in light microscopy image sequences*, in "10th International Symposium on Biomedical Imaging (ISBI)", IEEE, 2013, p. 468–471.

[55] M. PINOT, V. STEINER, B. DEHAPIOT, B.-K. YOO, F. CHESNEL, L. BLANCHOIN, C. KERVRANN, Z. GUEROUI.*Confinement induces actin flow in a meiotic cytoplasm*, in "Proceedings National Academy of Sciences USA", July 2012, vol. 109, n^o 29, p. 11705–11710 [*DOI :* 10.1073/PNAS.1121583109], http://hal.inria.fr/inserm-00717415.

[56] T. PÉCOT, C. KERVRANN, S. BARDIN, B. GOUD, J. SALAMERO.*Patch-based Markov models for event detection in fluorescence bioimaging*, in "Int. Conf. on Medical Image Computing and Computer Assisted Intervention (MICCAI'08)", New York City, USA, September 2008, vol. 2, p. 95–103.

[57] P. PÉREZ, J. VERMAAK, A. BLAKE.*Data fusion for visual tracking with particles*, in "Proc. IEEE", 2004, vol. 92, n^o 3, p. 495–513.

[58] I. SBALZARINI, P. KOUMOUTSAKOS.*Feature point tracking and trajectory analysis for video Imaging in cell biology*, in "J. Structural Biology", 2005, vol. 151, p. 182–195.

[59] I. SMAL, M. LOOG, W. NIESSEN, E. MEIJERING.*Quantitative comparison of spot detection methods in fluorescence microscopy*, in "IEEE Trans. Medical Imaging", 2010, vol. 29, n^o 2, p. 282–301.

[60] I. SMAL, E. MEIJERING, K. DRAEGESTEIN, N. GALJART, I. GRIGORIEV, A. AKHMANOVA, M. VAN ROYEN, A. HOUTSMULLER, W. NIESSEN.*Multiple object tracking in molecular bioimaging by Rao-Blackwellized marginal particle filtering,*, in "Medical Image Analysis", 2008, vol. 12, n^o 6, p. 764–777.

[61] D. THOMANN, D. RINES, P. SORGER, G. DANUSER.*Automatic fluorescent tag detection in 3D with super-resolution: application to the analysis of chromosome movement*, in "J. Microscopy", 2002, vol. 298, n^o Part 1, p. 49–64.

[62] Y. VARDI.*Network tomography: estimating source-destination traffic intensities from link data*, in "J. American Statistical Association", 1996, vol. 91, p. 365–377.

[63] T. WALTER, D. SHATTUCK, R. BALDOCK, M. BASTIN, A. CARPENTER, S. DUCE, J. ELLENBERG, A. FRASER, N. HAMILTON, S. PIEPER, M. RAGAN, J. SCHNEIDER, P. TOMANCAK, J. HERICHE.*Visualization of image data from cells to organisms*, in "Nature Methods", 2010, vol. 7, n^o 3 (Suppl), p. S26–41.

Project-Team SIROCCO

Analysis representation, compression and communication of visual data

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

IN PARTNERSHIP WITH:
CNRS

Université Rennes 1

RESEARCH CENTER
Rennes - Bretagne-Atlantique

THEME
Vision, perception and multimedia interpretation

Table of contents

Project-Team SIROCCO

Creation of the Project-Team: 2012 January 01

Keywords:

Computer Science and Digital Science:

5. - Interaction, multimedia and robotics
5.3. - Image processing and analysis
5.4. - Computer vision
5.9. - Signal processing

Other Research Topics and Application Domains:

6. - IT and telecom

1. Members

Research Scientists

Christine Guillemot [Team leader, Inria, Senior Researcher, HDR]
Thomas Maugey [Inria, Researcher]
Aline Roumy [Inria, 80%, Researcher]

Faculty Member

Olivier Le Meur [Univ. Rennes I, Associate Professor, HDR]

Technical Staff

Cedric Le Cam [Inria]
Mikael Le Pendu [Inria]
Xin Su [Inria]

PhD Students

Martin Alain [Technicolor, granted by CIFRE, until Mar 2016]
Jean Begaint [Technicolor, granted by CIFRE]
Pierre David [Inria, from Oct 2016]
Elian Dib [Inria, from Oct 2016]
Thierry Dumas [Inria]
Julio Cesar Ferreira [co-tutelle between the Univ. of Rennes 1 and the Univ. of Uberlandia, until July 2016]
David Gommelet [Cifre Ericsson, granted by CIFRE]
Fatma Hawary [Technicolor, granted by CIFRE]
Matthieu Hog [Technicolor, granted by CIFRE]
Hristina Hristova [Univ. Rennes I, also with FRVsense]
Dmitry Kuzovkin [Technicolor, granted by CIFRE, also with FRVSense]
Navid Mahmoudian Bidgoli [Inria, from Oct 2016]
Mira Rizkallah [Univ. Rennes I]
Maxime Rousselot [Harmonics, granted by CIFRE, also with FRVSense]

Post-Doctoral Fellows

Oriel Frigo [Inria, from Dec 2016]
Xiaoran Jiang [Inria]

Visiting Scientists

Reuben Farrugia [University of Malta, until Aug 2016]
Sheila S. Hemami [Univ. Rennes I, from May 2016 until Aug 2016]

Administrative Assistant

Huguette Bechu [Inria]

2. Overall Objectives

2.1. Introduction

The goal of the SIROCCO project-team is the design and development of algorithms and practical solutions in the areas of analysis, modelling, coding, and communication of images and video signals. The objective is to cover several inter-dependent algorithmic problems of the end-to-end transmission chain from the capturing, compression, transmission to the rendering of the visual data. The project-team activities are structured and organized around the following inter-dependent research axes:

- Analysis and modeling for compact representation and navigation [0] in large volumes of visual data [0]
- Rendering, inpainting and super-resolution of visual data
- Representation and compression of visual data
- Distributed processing and robust communication of visual data

Given the strong impact of standardization in the sector of networked multimedia, SIROCCO, in partnership with industrial companies, seeks to promote its results in standardization (MPEG). While aiming at generic approaches, some of the solutions developed are applied to practical problems in partnership with industry (Alcatel Lucent, Astrium, Orange labs., Technicolor, Thomson Video Networks) or in the framework of national projects (ANR-ARSSO, ANR-PERSEE). The application domains addressed by the project are networked visual applications taking into account their various requirements and needs in terms of compression, of resilience to channel noise and network adaptation, of advanced functionalities such as navigation, and of high quality rendering.

2.2. Analysis and modeling for compact representation

Analysis and modeling of the visual data are crucial steps for a number of video processing problems: navigation in 3D scenes, compression, loss concealment, denoising, inpainting, editing, content summarization and navigation. The focus is on the extraction of different cues such as scene geometry, edge, texture and motion, on the extraction of high-level features (GIST-like or epitomes), and on the study of computational models of visual attention, useful for different visual processing tasks. In relation to the above problems, the project-team considers various types of image modalities (medical and satellite images, natural 2D still and moving images, multi-view and multi-view plus depth video content).

2.3. Rendering, inpainting and super-resolution

This research axis addresses the problem of high quality reconstruction of various types of visual data after decoding. Depending on the application and the corresponding type of content (2D, 3D), various issues are being addressed. For example, to be able to render 3D scenes, depth information is associated with each view as a depth map, and transmitted in order to perform virtual view generation. Given one view with its depth information, depth image-based rendering techniques have the ability to render views in any other spatial positions. However, the issue of intermediate view reconstruction remains a difficult ill-posed problem. Most errors in the view synthesis are caused by incorrect geometry information, inaccurate camera parameters, and occlusions/disocclusions. Efficient inpainting techniques are necessary to restore disocclusions areas. Inpainting techniques are also required in transmission scenarios, where packet losses result in missing data in the video after decoding. The design of efficient mono-view and multi-view super-resolution methods is also part of the project-team objectives to improve the rendering quality, as well as to trade-off quality against transmission rate.

[0] By navigation we refer here to scene navigation by virtual view rendering, and to navigation across slices in volumic medical images.

[0] By visual data we refer to natural and medical images, videos, multi-view sequences as well as to visual cues or features extracted from video content.

2.4. Representation and compression of visual data

The objective is to develop algorithmic tools for constructing low-dimensional representations of multi-view video plus depth data, of 2D image and video data, of visual features and of their descriptors. Our approach goes from the design of specific algorithmic tools to the development of complete compression algorithms. The algorithmic problems that we address include data dimensionality reduction, the design of compact representations for multi-view plus depth video content which allow high quality 3D rendering, the design of sparse representation methods and of dictionary learning techniques. The sparsity of the representation indeed depends on how well the dictionary is adapted to the data at hand. The problem of dictionary learning for data-adaptive representations, that goes beyond the concatenation of a few traditional bases, has thus become a key issue which we address for further progress in the area.

Developing complete compression algorithms necessarily requires tackling visual processing topics beyond the issues of sparse data representation and dimensionality reduction. For example, problems of scalable, perceptual, and metadata-aided coding of 2D and 3D visual data, as well as of near lossless compression of medical image modalities (CT, MRI, virtual microscopy imaging) are tackled. Finally, methods for constructing rate-efficient feature digests allowing processing in lower-dimensional spaces, e.g. under stringent bandwidth constraints, also falls within the scope of this research axis.

2.5. Distributed processing and robust communication

The goal is to develop theoretical and practical solutions for robust image and video transmission over heterogeneous and time-varying networks. The first objective is to construct coding tools that can adapt to heterogeneous networks. This includes the design of (i) sensing modules to measure network characteristics, of (ii) robust coding techniques and of (iii) error concealment methods for compensating for missing data at the decoder when erasures occur during the transmission. The first objective is thus to develop sensing and modeling methods which can recognize, model and predict the packets loss/delay end-to-end behaviour. Given the estimated and predicted network conditions (e.g. Packet Error Rate (PER)), the objective is then to adapt the data coding, protection and transmission scheme. However, the reliability of the estimated PER impacts the performance of FEC schemes. We investigate the problem of constructing codes which would be robust to channel uncertainty, i.e. which would perform well not only on a specific channel but also “universally”, hence reducing the need for a feedback channel. This would be a significant advantage compared with rateless codes such as fountain codes which require a feedback channel. Another problem which we address is error concealment. This refers to the problem of estimating lost symbols from the received ones by exploiting spatial and/or temporal correlation within the video signal.

The availability of wireless camera sensors has also been spurring interest for a variety of applications ranging from scene interpretation, object tracking and security environment monitoring. In such camera sensor networks, communication energy and bandwidth are scarce resources, motivating the search for new distributed image processing and coding (Distributed Source Coding) solutions suitable for band and energy limited networking environments. In the past years, the team has developed a recognized expertise in the area of distributed source coding, which in theory allows for each sensor node to communicate losslessly at its conditional entropy rate without information exchange between the sensor nodes. However, distributed source coding (DSC) is still at the level of the proof of concept and many issues remain unresolved. The goal is thus to further address theoretical issues as the problem of modeling the correlation channel between sources, to further study the practicality of DSC in image coding and communication problems.

3. Research Program

3.1. Introduction

The research activities on analysis, compression and communication of visual data mostly rely on tools and formalisms from the areas of statistical image modelling, of signal processing, of coding and information theory. However, the objective of better exploiting the Human Visual System (HVS) properties in the above goals also pertains to the areas of perceptual modelling and cognitive science. Some of the proposed research axes are also based on scientific foundations of computer vision (e.g. multi-view modelling and coding). We have limited this section to some tools which are central to the proposed research axes, but the design of complete compression and communication solutions obviously rely on a large number of other results in the areas of motion analysis, transform design, entropy code design, etc which cannot be all described here.

3.2. Parameter Estimation and Inference

Bayesian estimation, Expectation-Maximization, stochastic modelling

Parameter estimation is at the core of the processing tools studied and developed in the team. Applications range from the prediction of missing data or future data, to extracting some information about the data in order to perform efficient compression. More precisely, the data are assumed to be generated by a given stochastic data model, which is partially known. The set of possible models translates the a priori knowledge we have on the data and the best model has to be selected in this set. When the set of models or equivalently the set of probability laws is indexed by a parameter (scalar or vectorial), the model is said parametric and the model selection resorts to estimating the parameter. Estimation algorithms are therefore widely used at the encoder to analyze the data. In order to achieve high compression rates, the parameters are usually not sent and the decoder has to jointly select the model (i.e. estimate the model parameters) and extract the information of interest.

3.3. Data Dimensionality Reduction

Manifolds, locally linear embedding, non-negative matrix factorization, principal component analysis

A fundamental problem in many data processing tasks (compression, classification, indexing) is to find a suitable representation of the data. It often aims at reducing the dimensionality of the input data so that tractable processing methods can then be applied. Well-known methods for data dimensionality reduction include principal component analysis (PCA) and independent component analysis (ICA). The methodologies which will be central to several proposed research problems will instead be based on sparse representations, on locally linear embedding (LLE) and on the “non negative matrix factorization” (NMF) framework.

The objective of *sparse representations* is to find a sparse approximation of a given input data. In theory, given $A \in \mathbb{R}^{m \times n}$, $m < n$, and $\mathbf{b} \in \mathbb{R}^m$ with $m << n$ and A is of full rank, one seeks the solution of $\min\{\|\mathbf{x}\|_0 \ : \ A\mathbf{x} = \mathbf{b}\}$, where $\|\mathbf{x}\|_0$ denotes the L_0 norm of x, i.e. the number of non-zero components in z. There exist many solutions x to $Ax = b$. The problem is to find the sparsest, the one for which x has the fewest non zero components. In practice, one actually seeks an approximate and thus even sparser solution which satisfies $\min\{\|\mathbf{x}\|_0 \ : \ \|A\mathbf{x} - \mathbf{b}\|_p \leq \rho\}$, for some $\rho \geq 0$, characterizing an admissible reconstruction error. The norm p is usually 2, but could be 1 or ∞ as well. Except for the exhaustive combinatorial approach, there is no known method to find the exact solution under general conditions on the dictionary A. Searching for this sparsest representation is hence unfeasible and both problems are computationally intractable. Pursuit algorithms have been introduced as heuristic methods which aim at finding approximate solutions to the above problem with tractable complexity.

Non negative matrix factorization (NMF) is a non-negative approximate data representation [0]. NMF aims at finding an approximate factorization of a non-negative input data matrix V into non-negative matrices W and H, where the columns of W can be seen as *basis vectors* and those of H as coefficients of the linear approximation of the input data. Unlike other linear representations like PCA and ICA, the non-negativity constraint makes the representation purely additive. Classical data representation methods like PCA or Vector Quantization (VQ) can be placed in an NMF framework, the differences arising from different constraints being placed on the W and H matrices. In VQ, each column of H is constrained to be unitary with only one non-zero coefficient which is equal to 1. In PCA, the columns of W are constrained to be orthonormal and the rows of H to be orthogonal to each other. These methods of data-dependent dimensionality reduction will be at the core of our visual data analysis and compression activities.

[0] D.D. Lee and H.S. Seung, “Algorithms for non-negative matrix factorization”, Nature 401, 6755, (Oct. 1999), pp. 788-791.

3.4. Perceptual Modelling

Saliency, visual attention, cognition

The human visual system (HVS) is not able to process all visual information of our visual field at once. To cope with this problem, our visual system must filter out irrelevant information and reduce redundant information. This feature of our visual system is driven by a selective sensing and analysis process. For instance, it is well known that the greatest visual acuity is provided by the fovea (center of the retina). Beyond this area, the acuity drops down with the eccentricity. Another example concerns the light that impinges on our retina. Only the visible light spectrum lying between 380 nm (violet) and 760 nm (red) is processed. To conclude on the selective sensing, it is important to mention that our sensitivity depends on a number of factors such as the spatial frequency, the orientation or the depth. These properties are modeled by a sensitivity function such as the Contrast Sensitivity Function (CSF).

Our capacity of analysis is also related to our visual attention. Visual attention which is closely linked to eye movement (note that this attention is called *overt* while the covert attention does not involve eye movement) allows us to focus our biological resources on a particular area. It can be controlled by both top-down (i.e. goal-directed, intention) and bottom-up (stimulus-driven, data-dependent) sources of information [0]. This detection is also influenced by prior knowledge about the environment of the scene [0]. Implicit assumptions related to prior knowledge or beliefs play an important role in our perception (see the example concerning the assumption that light comes from above-left). Our perception results from the combination of prior beliefs with data we gather from the environment. A Bayesian framework is an elegant solution to model these interactions [0]. We define a vector $\overrightarrow{v}_l$ of local measurements (contrast of color, orientation, etc.) and vector $\overrightarrow{v}_c$ of global and contextual features (global features, prior locations, type of the scene, etc.). The salient locations S for a spatial position $\overrightarrow{x}$ are then given by:

$$S(\overrightarrow{x}) = \frac{1}{p(\overrightarrow{v}_l\,|\,\overrightarrow{v}_c)} \times p(s, \overrightarrow{x}\,|\,\overrightarrow{v}_c) \tag{57}$$

The first term represents the bottom-up salience. It is based on a kind of contrast detection, following the assumption that rare image features are more salient than frequent ones. Most of existing computational models of visual attention rely on this term. However, different approaches exist to extract the local visual features as well as the global ones. The second term is the contextual priors. For instance, given a scene, it indicates which parts of the scene are likely the most salient.

3.5. Coding theory

OPTA limit (Optimum Performance Theoretically Attainable), Rate allocation, Rate-Distortion optimization, lossy coding, joint source-channel coding multiple description coding, channel modelization, oversampled frame expansions, error correcting codes.

Source coding and channel coding theory [0] is central to our compression and communication activities, in particular to the design of entropy codes and of error correcting codes. Another field in coding theory which has emerged in the context of sensor networks is Distributed Source Coding (DSC). It refers to the compression of correlated signals captured by different sensors which do not communicate between themselves. All the signals captured are compressed independently and transmitted to a central base station which has the capability to decode them jointly. DSC finds its foundation in the seminal Slepian-Wolf [0] (SW) and Wyner-Ziv [0]

---

[0] L. Itti and C. Koch, “Computational Modelling of Visual Attention” , Nature Reviews Neuroscience, Vol. 2, No. 3, pp. 194-203, 2001.

[0] J. Henderson, “Regarding scenes”, Directions in Psychological Science, vol. 16, pp. 219-222, 2007.

[0] L. Zhang, M. Tong, T. Marks, H. Shan, H. and G.W. Cottrell, “SUN: a Bayesian framework for saliency using natural statistics”, Journal of Vision, vol. 8, pp. 1-20, 2008.

[0] T. M. Cover and J. A. Thomas, Elements of Information Theory, Second Edition, July 2006.

[0] D. Slepian and J. K. Wolf, “Noiseless coding of correlated information sources.” IEEE Transactions on Information Theory, 19(4), pp. 471-480, July 1973.

[0] A. Wyner and J. Ziv, “The rate-distortion function for source coding ith side information at the decoder.” IEEE Transactions on Information Theory, pp. 1-10, January 1976.

(WZ) theorems. Let us consider two binary correlated sources X and Y. If the two coders communicate, it is well known from Shannon's theory that the minimum lossless rate for X and Y is given by the joint entropy $H(X, Y)$. Slepian and Wolf have established in 1973 that this lossless compression rate bound can be approached with a vanishing error probability for long sequences, even if the two sources are coded separately, provided that they are decoded jointly and that their correlation is known to both the encoder and the decoder.

In 1976, Wyner and Ziv considered the problem of coding of two correlated sources X and Y, with respect to a fidelity criterion. They have established the rate-distortion function $R*_{X|Y}(D)$ for the case where the side information Y is perfectly known to the decoder only. For a given target distortion D, $R*_{X|Y}(D)$ in general verifies $R_{X|Y}(D) \leq R*_{X|Y}(D) \leq R_X(D)$, where $R_{X|Y}(D)$ is the rate required to encode X if Y is available to both the encoder and the decoder, and R_X is the minimal rate for encoding X without SI. These results give achievable rate bounds, however the design of codes and practical solutions for compression and communication applications remain a widely open issue.

4. Application Domains

4.1. Introduction

The application domains addressed by the project are:

- Compression with advanced functionalities of various image modalities (including multi-view, medical images such as MRI, CT, WSI, or satellite images);
- Networked multimedia applications taking into account their various needs in terms of image and 2D and 3D video compression, or in terms of network adaptation (e.g., resilience to channel noise);
- Content editing and post-production.

4.2. Compression of emerging imaging modalities

Compression of images and of 2D video (including High Definition and Ultra High Definition) remains a widely-sought capability for a large number of applications. This is particularly true for mobile applications, as the need for wireless transmission capacity will significantly increase during the years to come. Hence, efficient compression tools are required to satisfy the trend towards mobile access to larger image resolutions and higher quality. A new impulse to research in video compression is also brought by the emergence of new formats beyond High Definition TV (HDTV) towards high dynamic range (higher bit depth, extended colorimetric space), super-resolution, formats for immersive displays allowing panoramic viewing and 3DTV.

Different video data formats and technologies are envisaged for interactive and immersive 3D video applications using omni-directional videos, stereoscopic or multi-view videos. The "omni-directional video" set-up refers to 360-degree view from one single viewpoint or spherical video. Stereoscopic video is composed of two-view videos, the right and left images of the scene which, when combined, can recreate the depth aspect of the scene. A multi-view video refers to multiple video sequences captured by multiple video cameras and possibly by depth cameras. Associated with a view synthesis method, a multi-view video allows the generation of virtual views of the scene from any viewpoint. This property can be used in a large diversity of applications, including Three-Dimensional TV (3DTV), and Free Viewpoint Video (FTV). The notion of "free viewpoint video" refers to the possibility for the user to choose an arbitrary viewpoint and/or view direction within a visual scene, creating an immersive environment. Multi-view video generates a huge amount of redundant data which need to be compressed for storage and transmission. In parallel, the advent of a variety of heterogeneous delivery infrastructures has given momentum to extensive work on optimizing the end-to-end delivery QoS (Quality of Service). This encompasses compression capability but also capability for adapting the compressed streams to varying network conditions. The scalability of the video content compressed representation and its robustness to transmission impairments are thus important features for seamless adaptation to varying network conditions and to terminal capabilities.

4.3. Networked visual applications

3D and Free Viewpoint TV: The emergence of multi-view auto-stereoscopic displays has spurred a recent interest for broadcast or Internet delivery of 3D video to the home. Multiview video, with the help of depth information on the scene, allows scene rendering on immersive stereo or auto-stereoscopic displays for 3DTV applications. It also allows visualizing the scene from any viewpoint, for scene navigation and free-viewpoint TV (FTV) applications. However, the large volumes of data associated to multi-view video plus depth content raise new challenges in terms of compression and communication.

Internet and mobile video: Broadband fixed (ADSL, ADSL2+) and mobile access networks with different radio access technologies (RAT) (e.g. 3G/4G, GERAN, UTRAN, DVB-H), have enabled not only IPTV and Internet TV but also the emergence of mobile TV and mobile devices with internet capability. A major challenge for next internet TV or internet video remains to be able to deliver the increasing variety of media (including more and more bandwidth demanding media) with a sufficient end-to-end QoS (Quality of Service) and QoE (Quality of Experience).

Mobile video retrieval: The Internet has changed the ways of interacting with content. The user is shifting its media consumption from a passive to a more interactive mode, from linear broadcast (TV) to on demand content (YouTubes, iTunes, VoD), and to user-generated, searching for relevant, personalized content. New mobility and ubiquitous usage has also emerged. The increased power of mobile devices is making content search and retrieval applications using mobile phones possible. Quick access to content in mobile environments with restricted bandwidth resources will benefit from rate-efficient feature extraction and description.

Wireless multi-camera vision systems: Our activities on scene modelling, on rate-efficient feature description, distributed coding and compressed sensing should also lead to algorithmic building blocks relevant for wireless multi-camera vision systems, for applications such as visual surveillance and security.

4.4. Editing and post-production

Video editing and post-production are critical aspects in the audio-visual production process. Increased ways of “consuming” video content also highlight the need for content repurposing as well as for higher interaction and editing capabilities. Content captured at very high resolutions may need to be repurposed in order to be adapted to the requirements of actual users, to the transmission channel or to the terminal. Content repurposing encompasses format conversion (retargeting), content summarization, and content editing. This processing requires powerful methods for extracting condensed video representations as well as powerful inpainting techniques. By providing advanced models, advanced video processing and image analysis tools, more visual effects, with more realism become possible. Other applications such as video annotation/retrieval, video restoration/stabilization, augmented reality, can also benefit from the proposed research.

5. Highlights of the Year

5.1. Highlights of the Year

5.1.1. Awards

C. Guillemot has been granted an ERC advanced grant for a project on computational light fields imaging.

6. New Software and Platforms

6.1. Salient object extraction

FUNCTIONAL DESCRIPTION This software detects salient object in an input picture in an automatic manner. The detection is based on super-pixel segmentation and contrast of histogram. This software is dedicated to people working in image processing and post production.

- Participants: Zhi Liu and Olivier Le Meur
- Contact: Olivier Le Meur

6.2. VideoInpainting

KEYWORDS: Video Inpainting - Motion informations - Loss concealment - BMFI (Bilinear Motion Field Interpolation)

SCIENTIFIC DESCRIPTION From an input binary mask and a source picture, the software performs an examplar-based inpainting. The method is based on the combination of multiple inpainting applied on a low resolution of the input picture. Once the combination has been done, a single-image super-resolution method is applied to recover the details and the high frequency in the inpainted areas. The developments have been pursued in 2014, in particular by introducing a Poisson blending step in order to improve the visual quality of the inpainted video. This software is dedicated to people working in image processing and post production.

- Participants: Ronan Le Boulch and Olivier Le Meur
- Contact: Olivier Le Meur

6.3. Visual Fixation Analysis

SCIENTIFIC DESCRIPTION From a set of fixation data and a picture, the software called Visual Fixation Analysis extracts from the input data a number of features (fixation duration, saccade length, orientation of saccade...) and computes a human saliency map. The software can also be used to assess the degree of similarity between a ground truth (eye fixation data) and a predicted saliency map. This software is dedicated to people working in cognitive science and computer vision.

- Participants: Olivier Le Meur and Thierry Baccino
- Partner: Université de Rennes 1
- Contact: Olivier Le Meur

6.4. Saccadic model

SCIENTIFIC DESCRIPTION The software called Scanpath Prediction aims at predicting the visual scanpath of an observer. The visual scanpath is a set of fixation points. The computational model is based on bottom-up saliency maps, viewing tendencies (that have been learned from eye tracking datasets) and inhibition-of-return. A presentation of this model is available on the following link: http://fr.slideshare.net/OlivierLeMeur/saccadic-model-of-eye-movements-for-freeviewing-condition. This software is dedicated to people working in computer science, computer vision and cognitive science. This software is being registered at the APP (Agence de Protection des Programmes) under the number IDDN.FR.001.240029.000.S.P.2016.000.10000.

- Participants: Olivier Le Meur
- Partner: Université de Rennes 1
- Contact: Olivier Le Meur

6.5. ADT-ATep

The scientific and industrial community is nowadays exploring new multimedia applications using 3D data (beyond stereoscopy). In particular, Free Viewpoint Television (FTV) has attracted much attention in the recent years. In those systems, the user can choose in real time the view angle from which he wants to observe the scene. Despite the great interest for FTV, the lack of realistic and ambitious datasets penalizes the research effort. The acquisition of such sequences is very costly in terms of hardware and working effort, which explains why no multi-view videos suitable for FTV has been made available yet.

A project founded by Inriahub has recently started in the SIROCCO team. Called ATeP for “Acquisition, Traitement et Partage" (Acquisition, Processing and Sharing), it targets the acquisition of such dataset. Another interesting aspect of this project is that the acquisition system relies on omnidirectional cameras. The dataset will thus interest all the industries and scientists currently working on the development of efficient processing and coding tools for 360 videos.

- Participants: Cédric Le Cam, Thomas Maugey
- Partner: Inria
- Contact: Thomas Maugey

7. New Results

7.1. Analysis and modeling for compact representation

3D modelling, multi-view plus depth videos, light-fields, 3D meshes, epitomes, image-based rendering, inpainting, view synthesis

7.1.1. Visual attention

Participant: Olivier Le Meur.

Visual attention is the mechanism allowing to focus our visual processing resources on behaviorally relevant visual information. Two kinds of visual attention exist: one involves eye movements (overt orienting) whereas the other occurs without eye movements (covert orienting). Our research activities deal with the understanding and modeling of overt attention.

Saccadic model: Previous research showed the existence of systematic tendencies in viewing behavior during scene exploration. For instance, saccades are known to follow a positively skewed, long-tailed distribution, and to be more frequently initiated in the horizontal or vertical directions. In 2016, we investigated the fact that these viewing biases are not universal, but are modulated by the semantic visual category of the stimulus. We showed that the joint distribution of saccade amplitudes and orientations significantly varies from one visual category to another. These joint distributions turn out to be, in addition, spatially variant within the scene frame. We demonstrated that a saliency model based on this better understanding of viewing behavioral biases and blind to any visual information outperforms well-established saliency models. We also proposed an extension of the saccadic model developed in 2015. The improvement consists in accounting for spatially-variant and context-dependent viewing biases. This model outperforms state-of-the-art saliency models, and provides scanpaths in close agreement with human behavior.

Inference of age from eye movements: We have presented evidence that information derived from eye gaze can be used to infer observers' age. From simple features extracted from the sequence of fixations and saccades, we predict the age of an observer. To reach this objective, we used the eye data from 101 observers split in 4 age groups (adults, 6-10 year-old, 4-6 year-old. and 2 year-old) to train a computational model. Participant's eye movements were monitored while participants were instructed to explore color pictures taken from children books for 10 seconds. The analysis of eye gaze provided evidence of age-related differences in viewing patterns. Fixation durations decreased with age while saccades turned out to be shorter when comparing children with adults. We combine several features, such as fixation durations, saccade amplitudes, and learn a direct mapping from those features to age using Gentle AdaBoost classifiers. Experimental results show that the proposed method succeeds in predicting reasonably well the observer's age.

7.1.2. Graph structure in the rays space for fast light fields segmentation

Participants: Christine Guillemot, Matthieu Hog.

In collaboration with Technicolor (Neus Sabater), we have introduced a novel graph representation for interactive light field segmentation using Markov Random Field (MRF). The greatest barrier to the adoption of MRF for light field processing is the large volume of input data. The proposed graph structure exploits the redundancy in the ray space in order to reduce the graph size, decreasing the running time of MRF-based optimisation tasks. The concepts of free rays and ray bundles with corresponding neighbourhood relationships are defined to construct the simplified graph-based light field representation. We have then developed a light field interactive segmentation algorithm using graph-cuts based on such ray space graph structure, that guarantees the segmentation consistency across all views. Our experiments with several datasets show results that are very close to the ground truth, competing with state of the art light field segmentation methods in terms of accuracy and with a significantly lower complexity. They also show that our method performs well on both densely and sparsely sampled light fields [18] (see Figure 1).

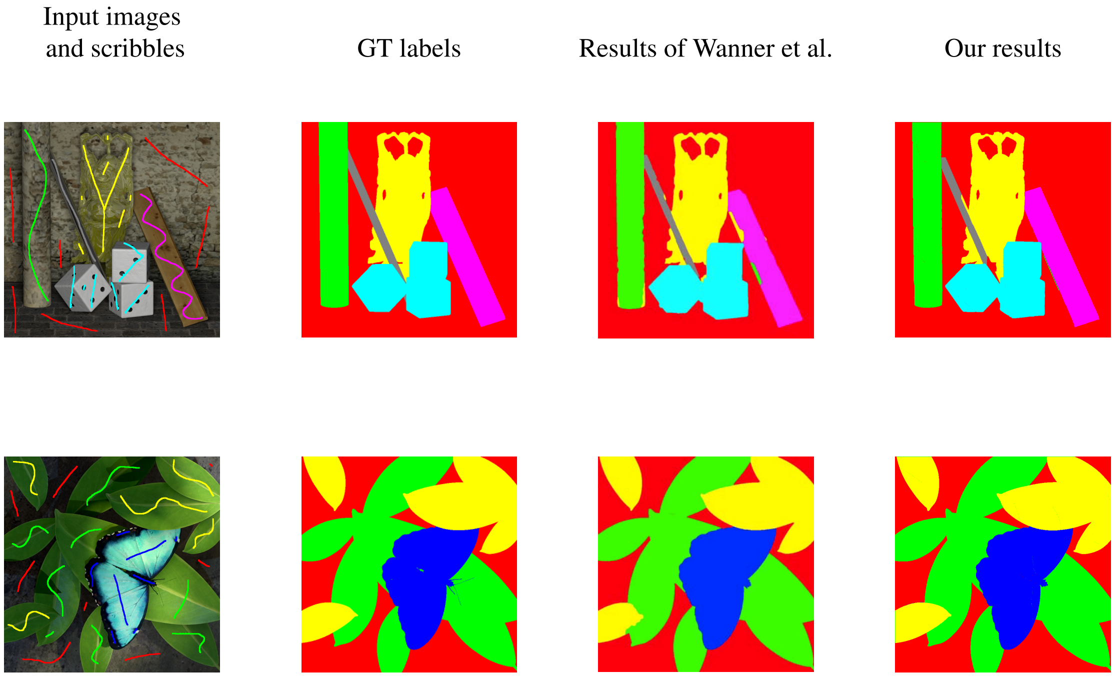


Figure 1. Light-field segmentation results obtained with synthetic light-fields. From left to right, we show, the input central view with scribbles, the ground truth labelling, the results of Wanner et al. and our results.

7.2. Rendering, inpainting and super-resolution

image-based rendering, inpainting, view synthesis, super-resolution

7.2.1. Joint color and gradient transfer through Multivariate Generalized Gaussian Distribution

Participants: Hristina Hristova, Olivier Le Meur.

Multivariate generalized Gaussian distributions (MGGDs) have aroused a great interest in the image processing community thanks to their ability to describe accurately various image features, such as image gradient fields, wavelet coefficients, etc. However, so far their applicability has been limited by the lack of a transformation between two of these parametric distributions. In collaboration with FRVSense (Rémi Cozot and Kadi Bouatouch), we have proposed a novel transformation between MGGDs, consisting of an optimal transportation of the second-order statistics and a stochastic-based shape parameter transformation. We employ the proposed transformation in both color and gradient transfers between images. We have also proposed a new simultaneous transfer of color and gradient.

7.2.2. High-Dynamic-Range Image Recovery from Flash and Non-Flash Image Pairs

Participants: Hristina Hristova, Olivier Le Meur.

In 2016, in collaboration with FRVSense (Rémi Cozot and Kadi Bouatouch), we have proposed a novel method for creating High Dynamic Range (HDR) images from only two images - flash and non-flash images. The proposed method consists of two main steps, namely brightness gamma correction and bi-local chromatic adaptation transform (CAT). First, the brightness gamma correction performs series of increases and decreases of the brightness of the non-flash image and that way yields multiple images with various exposure values. Second, a proposed CAT method, called bi-local CAT enhances the quality of the computed images, by recovering details in the under-/over-exposed regions, using detail information from the flash image. The final multiple exposure images are then merged together to compute an HDR image. Evaluation shows that our HDR images, obtained by using only two LDR images, are close to HDR images, obtained by combining five manually taken multi-exposure images. The proposed method does not require the usage of a tripod and it is suitable for images of non-still objects, such as people, candle flames, etc. Figure 2 illustrates some results of the proposed method. The HDR-VDP-2 color-coded map (right-most image) shows the main luminance differences (the red areas) between our HDR result and the real HDR image. Snippets (a) and (b) show that the proposed method sharpens fine details, e.g. the net on the lamp. The net on the lamp of the real HDR image is blurry, due to a movement in the real multi-exposure images.

Figure 2. HDR image recovery from two input images, i.e. flash and non-flash images. Our HDR result and the real HDR image are tone-mapped for visualization on an LDR display.

7.2.3. Depth inpainting

Participant: Olivier Le Meur.

To tackle the disocclusion inpainting of RGB-D images appearing when synthesizing new views of a scene by changing its viewpoint, in collaboration with Pierre Buyssens from the Greyc laboratory from the Caen University, we have developed a new examplar-based inpainting method of depth map. The proposed method is based on two main components. First, a novel algorithm to perform the depth-map disocclusion inpainting has been proposed. In particular, this intuitive approach is able to recover the lost structures of the objects and to inpaint the depth-map in a geometrically plausible manner. Then, a depth-guided patch-based inpainting method has been defined in order to fill-in the color image. Depth information coming from the reconstructed depth-map is added to each key step of the classical patch-based algorithm from Criminisi et al. in an intuitive manner. Relevant comparisons to state-of-the-art inpainting methods for the disocclusion inpainting of both depth and color images have illustrated the effectiveness of the proposed algorithms.

7.2.4. Super-resolution and inpainting for face recognition

Participants: Reuben Farrugia, Christine Guillemot.

Most face super-resolution methods assume that low- and high-resolution manifolds have similar local geometrical structure, hence learn local models on the low-resolution manifold (e.g. sparse or locally linear embedding models), which are then applied on the high-resolution manifold. However, the low- resolution manifold is distorted by the one-to-many relationship between low- and high- resolution patches.

We have developed a method which learns linear models based on the local geometrical structure on the high-resolution manifold rather than on the low-resolution manifold. For this, in a first step, the low-resolution patch is used to derive a globally optimal estimate of the high-resolution patch. The approximated solution is shown to be close in Euclidean space to the ground-truth but is generally smooth and lacks the texture details needed by state-of-the-art face recognizers. Unlike existing methods, the sparse support that best estimates the first approximated solution is found on the high-resolution manifold. The derived support is then used to extract the atoms from the coupled dictionaries that are most suitable to learn an upscaling function between the low- and high- resolution patches.

The proposed solution has also been extended to compute face super-resolution of non-frontal images. Experimental results show that the proposed method out- performs six face super-resolution and a state-of-the-art cross- resolution face recognition method. These results also reveal that the recognition and quality are significantly affected by the method used for stitching all super-resolved patches together, where quilting was found to better preserve the texture details which helps to achieve higher recognition rates. The proposed method was shown to be able to super-resolve facial images from the IARPA Janus Benchmark A (IJB-A) dataset which considers a wide range of poses and orientations.

A method has also been developed to inpaint occluded facial regions with unconstrained pose and orientation. This approach first warps the facial region onto a reference model to synthesize a frontal view [15]. A modified Robust Principal Component Analysis (RPCA) approach is then used to suppress warping errors. It then uses a novel local patch-based face inpainting algorithm which hallucinates missing pixels using a dictionary of face images which are pre-aligned to the same reference model. The hallucinated region is then warped back onto the original image to restore missing pixels. Experimental results on synthetic occlusions demonstrate that the proposed face inpainting method has the best performance achieving PSNR gains of up to 0.74dB over the second-best method. Moreover, experiments on the COFW dataset and a number of real-world images show that the proposed method successfully restores occluded facial regions in the wild even for Closed-Circuit Television (CCTV) quality images.

7.2.5. *Light-field inpainting*

Participants: Christine Guillemot, Xiaoran Jiang, Mikael Le Pendu.

Building up on the advances in low rank matrix completion, we have developed a novel method for propagating the inpainting of the central view of a light field to all the other views. After generating a set of warped versions of the inpainted central view with random homographies, both the original light field views and the warped ones are vectorized and concatenated into a matrix. Because of the redundancy between the views, the matrix satisfies a low rank assumption enabling us to fill the region to inpaint with low rank matrix completion. To this end, a new matrix completion algorithm, better suited to the inpainting application than existing methods, has also been developed. Unlike most of the existing light field inpainting algorithms, our method does not require any depth prior. Another interesting feature of the low rank approach is its ability to cope with color and illumination variation between the input views of the light field (see Fig.3. As it can be seen in Figure 3, the proposed method yields inpainting consistency across views.

(a) (b) (c) (d)

Figure 3. Illustration of our inpainting propagation method : (a) Original central view. (b) Inpainted central view. (c) Another view of the light field inpainted with a state-of-the-art 2D image inpainting method. (d) Propagated inpainting from central view to a different view with the developed low rank method.

7.3. Representation and compression of large volumes of visual data

Sparse representations, data dimensionality reduction, compression, scalability, perceptual coding, rate-distortion theory

7.3.1. Graph-based multi-view video representation

Participants: Christine Guillemot, Thomas Maugey, Mira Rizkallah, Xin Su.

One of the main open questions in multiview data processing is the design of representation methods for multiview data, where the challenge is to describe the scene content in a compact form that is robust to lossy data compression. Many approaches have been studied in the literature, such as the multiview and multiview plus depth formats, point clouds or mesh-based techniques. All these representations contain two types of data: i) the color or luminance information, which is classically described by 2D images; ii) the geometry information that describes the scene 3D characteristics, represented by 3D coordinates, depth maps or disparity vectors. Effective representation, coding and processing of multiview data partly rely on a proper representation of the geometry information. The multiview plus depth (MVD) format has become very popular in recent years for 3D data representation. However, this format induces very large volumes of data, hence the need for efficient compression schemes. On the other hand, lossy compression of depth information in general leads to annoying rendering artefacts especially along the contours of objects in the scene. Instead of lossy compression of depth maps, we consider the lossless transmission of a geometry representation that captures only the information needed for the required view reconstructions.

The goal is thus to develop a Graph-Based Representation (GBR) for geometry information, where the geometry of the scene is represented as connections between corresponding pixels in different views. In this representation, two connected pixels are neighboring points in the 3D scene. The graph connections are derived from dense disparity maps and provide just enough geometry information to predict pixels in all the views that have to be synthesized. GBR drastically simplifies the geometry information to the bare minimum required for view prediction. This "task-aware" geometry simplification allows us to control the view prediction accuracy before coding compared to baseline depth compression methods. In 2015, we have first considered multi-view configurations, in which cameras are parallel.

In 2016, we have developed the extension of GBR to complex camera configurations. In [21], Xin Su has implemented a generalized Graph-Based Representation handling two views with complex translations and rotations between them (Fig. 4). The proposed approach uses the epipolar segments to have a row-wise description of the geometry that is as simple as for rectified views. This generalized GBR has been further extended to handle multiple views and scalable description of the geometry, *i.e.*, a geometry data that is coded as a function of the user navigation among the views.

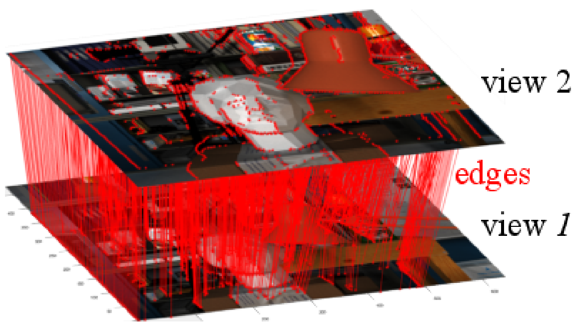


Figure 4. The proposed GBR (i) provides edges describing the geometry information and (ii) link pixels that are neighbors in the 3D scene.

The graph described above links neighboring pixels in the 3D scene as 3D meshes do. This meaningful structure might be used to code the color pixels lying on it. This can be done thanks to the new processing tools developed for signals lying on graphs. These tools rely however on covariance models that are assumed to be suited for the processed data. The PhD work of Mira Rizkallah is currently focussing on the effect of errors in the correlation models on the efficiency of the graph-based transforms.

7.3.2. Sparse and low rank approximation of light fields

Participants: Christine Guillemot, Xiaoran Jiang, Mikael Le Pendu.

We have studied the problem of low rank approximation of light fields for compression. A homography-based approximation method has been proposed which jointly searches for homographies to align the different views of the light field together with the low rank approximation matrices. We have first considered a global homography per view and shown that depending on the variance of the disparity across views, the global homography is not sufficient to well-align the entire images. In a second step, we have thus considered multiple homographies, one per region, the region being extracted using depth information. We have first shown the benefit of the joint optimization of the homographies together with the low-rank approximation. The resulting compact representation compressed using HEVC yields compression performance significantly superior to those obtained by directly applying HEVC on the light field views re-structured as a video sequence.

7.3.3. Deep learning, autoencoders and neural networks for sparse representation and compression

Participants: Thierry Dumas, Christine Guillemot, Aline Roumy.

Deep learning is a novel research area that attempts to extract high level abstractions from data by using a graph with multiple layers. One could therefore expect that deep learning might allow efficient image compression based on these high level features. However, deep learning, as classical machine learning, consists in two phases: (i) build a graph that can make a good representation of the data (i.e. find an architecture usually made with neural nets), and (ii) learn the parameters of this architecture from large-scale data. As a consequence, neural nets are well suited for a specific task (text or image recognition) and require one training per task. The difficulty to apply machine learning approach to image compression is that it is important to deal with a large variety of patches, and with also various compression rates. To test the ability of neural networks to compress images, we studied shallow sparse autoencoders (AE) for image compression in [14]. A performance analysis in terms of rate-distortion trade-off and complexity is conducted, comparing sparse AEs with LARS-Lasso, Coordinate Descent (CoD) and Orthogonal Matching Pursuit (OMP). A Winner Take All Auto-encoder (WTA AE) is proposed where image patches compete with one another when computing their sparse representation. This allows to spread the sparsity constraint on the whole image. Since the learning is made for this WTA AE, the neural network also learns to deal with various patches, which helps building a general-purpose AE. Finally, we showed that, WTA AE achieves the best rate-distortion trade-off, is robust to quantization noise and it is less complex than LARS-Lasso, CoD and OMP.

7.3.4. Data geometry aware local basis selection

Participants: Julio Cesar Ferreira, Christine Guillemot.

Local learning of sparse image models has proven to be very effective to solve a variety of inverse problems in many computer vision applications. To learn such models, the data samples are often clustered using the K-means algorithm with the Euclidean distance as a dissimilarity metric. However, the Euclidean distance may not always be a good dissimilarity measure for comparing data samples lying on a manifold.

In 2015, we have developed, in collaboration with Elif Vural (now Prof. at METU in Ankara, former postdoc in the team), two algorithms for determining a local subset of training samples from which a good local model can be computed for reconstructing a given input test sample, where we take into account the underlying geometry of the data. The first algorithm, called Adaptive Geometry-driven Nearest Neighbor search (AGNN), is an adaptive scheme which can be seen as an out-of-sample extension of the replicator graph clustering method for local model learning. The second method, called Geometry-driven Overlapping Clusters (GOC), is a less complex nonadaptive alternative for training subset selection. The AGNN and GOC methods have been evaluated in image super-resolution, deblurring and denoising applications and shown to outperform spectral clustering, soft clustering, and geodesic distance based subset selection in most settings. The selected patches are used for learning good local bases using the traditional PCA method. PCA is considered an efficient tool to recover the tangent space of the patch manifold when the manifold is sufficiently regular.

However, when the patch manifold has high curvature, which is observed to be the case for images with high frequencies, PCA may not be suitable. It can be seen in Figure 5 that the PCA basis with respect to a manifold fails to approximate the tangent space as the manifold bends over itself. In other words, PCA basis is not adapted when the curvature is too high. On the other hand, it can be seen in Figure 5 that a union of subspaces with respect to a manifold might generate a local model that yields a more efficient local representation of data.

In 2016, we have proposed a strategy to choose between these two kinds of bases locally depending on the local data geometry. This function is defined as the variability of the tangent space in each cluster.

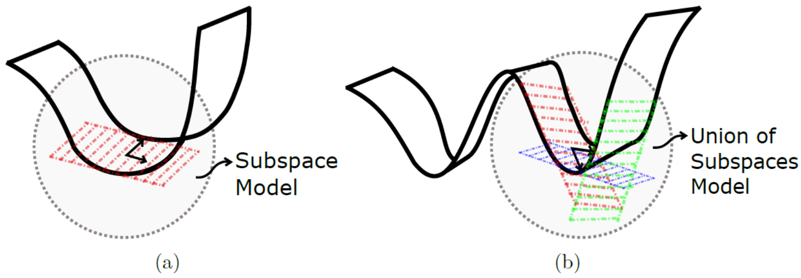


Figure 5. Subspaces computed with data sampled from a neighborhood on a manifold; (a): PCA basis which fails to approximate the subspace as the manifold curvature is too high; (b): union of subspaces generating a local model more coherent with the manifold geometry.

7.3.5. Rate-distortion optimized tone curves for HDR video compression

Participants: David Gommelet, Christine Guillemot, Aline Roumy.

High Dynamic Range (HDR) images contain more intensity levels than traditional image formats. Instead of 8 or 10 bit integers, floating point values requiring much higher precision are used to represent the pixel data. These data thus need specific compression algorithms. In collaboration with Ericsson [17], we have developed a novel compression algorithm that allows compatibility with the existing Low Dynamic Range (LDR) broadcast architecture in terms of display, compression algorithm and datarate, while delivering full HDR data to the users equipped with HDR display. The developed algorithm is thus a scalable video compression offering a base layer that corresponds to the LDR data and an enhancement layer, which together with the base layer corresponds to the HDR data. The novelty of the approach relies on the optimization of a mapping called Tone Mapping Operator (TMO) that maps efficiently the HDR data to the LDR data. The optimization has been carried out in a rate-distortion sense: the distortion of the HDR data is minimized under the constraint of minimum sum datarate (for the base and enhancement layer), while offering LDR data that are closed to some "aesthetic" a priori. Taking into account the aesthetic of the scene in video compression is indeed novel, since video compression is traditionally optimized to deliver the smallest distortion with the input data at the minimum datarate.

7.3.6. Cloud-based image compression

Participants: Jean Begaint, Christine Guillemot.

The emergence of cloud applications and web services has led to an increasing use of online resources. Image processing applications can benefit from this vast storage and distribution capacity. In collaboration with Technicolor, we investigate the use of this mass of redundant data to enhance image compression schemes. A region-based registration algorithm has been developped to capture complex deformations between two images. The registration method is then used to exploit both global and local correspondences between pairs of images of the same scene. The region-based registration yields a better prediction (hence reduced prediction errors, see Fig.6) which in turn yields a significant rate-distortion performance gain compared to current image coding solutions.

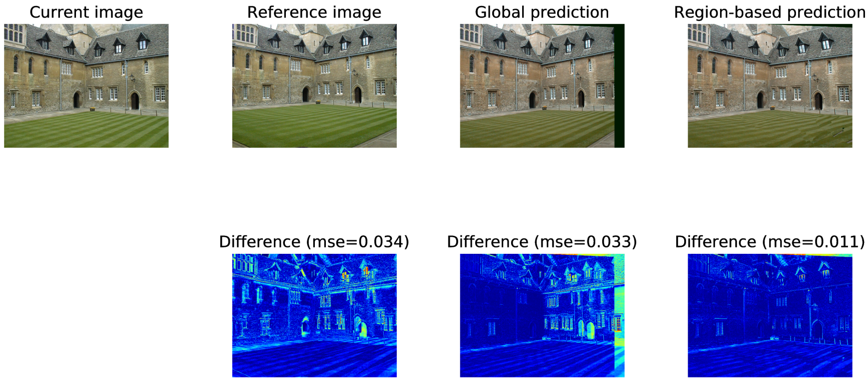


Figure 6. Image registration with global and region-based homographies and corresponding prediction error.

7.4. Distributed processing and robust communication

Information theory, stochastic modelling, robust detection, maximum likelihood estimation, generalized likelihood ratio test, error and erasure resilient coding and decoding, multiple description coding, Slepian-Wolf coding, Wyner-Ziv coding, information theory, MAC channels

7.4.1. Interactive Coding for Navigation in 3D scenes (ICON 3D)

Participants: Thomas Maugey, Aline Roumy.

In order to have performing FTV systems, the data transmission has to take into account the interactivity of the user, *i.e.,* the viewpoint that is requested. In other words, a FTV system transmits to the visualisation support only what needs to be updated when a user changes its viewpoint angle (*i.e.*, the new information appearing in its vision field). The Sirocco has recently proposed some promising work using channel coding for interactive data coding. This coding scheme focusses on multi-view plus depth format only. In order to extend this approach to other formats, we have started a collaboration with the I3S laboratory in Nice, expert in 3D mesh compression.

The project ICON 3D funded by the GdR-Isis will be divided into two parts. First, we will study and develop new geometry prediction algorithms for surface meshes. Given a part of a mesh, the prediction algorithm should be able to estimate a neighboring mesh subset corresponding to the one newly visible after user viewpoint angle change (Fig. 7). The prediction error will be characterized. Then, we will study the channel coding method that should be developed to correct this error.

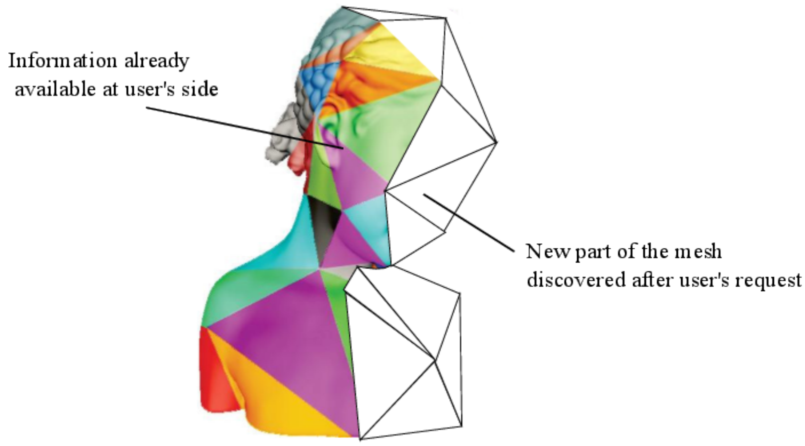


Figure 7. When a user changes his viewpoint angle, he discovers new part of the mesh that has to be transmitted.

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Grants with Industry

8.1.1. Consulting contract with Enensys technologies

Participant: Aline Roumy.

- Title : Matrix inversion for video streaming.
- Research axis : REF AT 7.4. Distributed processing and robust communication
- Partners : Enensys, Inria-Rennes.
- Funding : Enensys.
- Period : Apr. 2016 - May 2016.

This contract with Enensys technologies aimed at studying solutions for reducing the complexity of matrix inversion used for encoding data in the context of video streaming. First a bibliographical study has been carried out related to the problem of matrix inversion in a finite field, then a novel solution has been proposed together with some recommendations regarding the algorithmic implementation.

8.1.2. Google faculty research award

Participants: Christine Guillemot, Xiaoran Jiang, Mikael Le Pendu.

- Title : Light fields low rank and sparse approximation
- Research axis : 7.3.2
- Partners : Inria-Rennes.
- Funding : Google.
- Period : Oct.2015-Sept.2016.

The goal of the project was to study low-rank and sparse approximation models for light fields compression. A homography-based low-rank approximation has been developed showing significant PSNR-rate gains compared to a direct encoding of light field views with HEVC-inter coding.

8.1.3. CIFRE contract with Envivio/ Ericsson on LDR compatible HDR video coding

Participants: Christine Guillemot, David Gommelet, Aline Roumy.

- Title : LDR-compatible coding of HDR video signals.
- Research axis : § 7.3.5.
- Partners : Envivio.
- Funding : Cifre Envivio.
- Period : Oct.2014-Sept.2017.

The goal of this Cifre contract is to design solutions for LDR-compatible coding of HDR videos. This involves the study of rate-distortion optimized tone mapping operators taking into account constraints of temporal coherency to avoid the temporal flickering which results from a direct frame-by-frame application of classical tone mapping operators. The goal is also to design a coding architecture which will build upon these operators, integrating coding tools tailored to the statistics of the HDR refinement signals.

8.1.4. CIFRE contract with Harmonic on image analysis for HDR video compression

Participants: Maxime Rousselot, Olivier Le Meur.

- Title : image and video analysis for HDR video compression
- Partners : Harmonic, Univ. Rennes 1
- Funding: Harmonic, ANRT
- Period: April 2016-April 2019

This project (in collaboration with Rémi Cozot, FRVSense) aims to investigate two main axes. First, we want to assess whether the representation of High Dynamic Range signal has an impact on the coding efficiency. We will focus mainly on the Hybrid Log-Gamma (HLG) and Perceptual Quantizer (PQ) OETF (Opto-Electronic Transfer Function)approaches. The former defines a nonlinear transfer function which is display-independent and able to produce high quality images without compromising the director's artistic intent. The latter approach is based on Just Noticeable Difference curve. If it turns out that this representation has an impact, the coding strategy should be adjusted with respect to the representation. In addition, specific preprocessing tools will be defined to deal with the limitations of PQ and HLG approaches.

8.1.5. CIFRE contract with Technicolor on image collection analysis

Participants: Dmitry Kuzovkin, Olivier Le Meur.

- Title : Spatiotemporal retargeting and recomposition based on artistic rules
- Partners : Technicolor, Univ. Rennes 1
- Funding: Technicolor, ANRT
- Period: Nov. 2015 – Nov. 2018

The goal of the project (in collaboration with Rémi Cozot, FRVSense) is to take advantage of the huge quantities of image and video data currently available - captured by both amateur and professional users - as well as the multiple copies of each scene that users often capture, to improve the aesthetic appeal of content. Additionally, given Technicolor's unique position, we propose to take advantage of insights as well as content from professional artists and colorists to learn how different content types can be enhanced.

8.1.6. CIFRE contract with Technicolor on light fields editing

Participants: Christine Guillemot, Matthieu Hog.

- Title : Light fields editing
- Research axis : 7.1.2
- Partners : Technicolor, Inria-Rennes.
- Funding : Technicolor, ANRT.
- Period : Oct.2015-Sept.2018.

Editing is quite common with classical imaging. Now, if we want light-fields cameras to be in the future as common as traditional cameras, this functionality should also be enabled with light-fields. The goal of the PhD is to develop methods for light-field editing, and the work in 2016 has focused on the design of fast semi-supervised segmentation algorithms with coherence constraints across sub-aperture images (see 7.1.2).

8.1.7. CIFRE contract with Technicolor on light fields compressed representation

Participants: Christine Guillemot, Fatma Hawary.

- Title : Light fields compressed representation
- Partners : Technicolor, Inria-Rennes.
- Funding : Technicolor, ANRT.
- Period : Feb.2016-Jan.2019.

The goal of this PhD is to study reconstruction algorithms from compressed measurements based on the assumption of sparsity in the Fourier domain. The goal is to apply these algorithms to scalable compression of light fields.

8.1.8. CIFRE contract with Technicolor on cloud-based image compression

Participants: Jean Begaint, Christine Guillemot.

- Title : Cloud-based image compression
- Research axis : 7.3.6
- Partners : Technicolor, Inria-Rennes.
- Funding : Technicolor, ANRT.
- Period : Nov.2015-Oct.2018.

The goal of this Cifre contract is to develop a novel image compression scheme exploiting similarity between images in a cloud. The objective will therefore be to develop rate-distortion optimized affine or homographic estimation and compensation methods which will allow us to construct prediction schemes and learn adapted bases from most similar images retrieved by image descriptors. One issue to be addressed is the rate-distortion trade-off induced by the need for transmitting image descriptors.

9. Partnerships and Cooperations

9.1. Regional Initiatives

9.1.1. CominLabs/InterCom project

Participants: Aline Roumy, Thomas Maugey.

- Title : Interactive Communication (INTERCOM): Massive random access to subsets of compressed correlated data .
- Research axis : 7.4.1
- Partners : Inria-Rennes (Sirocco team and i4S team); LabSTICC, Telecom Bretagne, Signal & Communications Department; External partner: Kieffer L2S, CentraleSupelec, Univ. Paris Sud.
- Funding : Labex CominLabs.
- Period : Oct. 2016 - Nov. 2019.

This project aims to develop novel compression techniques allowing massive random access to large databases. Indeed, we consider a database that is so large that, to be stored on a single server, the data have to be compressed efficiently, meaning that the redundancy/correlation between the data have to be exploited. The dataset is then stored on a server and made available to users that may want to access only a subset of the data. Such a request for a subset of the data is indeed random, since the choice of the subset is user-dependent. Finally, massive requests are made, meaning that, upon request, the server can only perform low complexity operations (such as bit extraction but no decompression/compression). Algorithms for two emerging applications of this problem will be developed: Free-viewpoint Television (FTV) and massive requests to a database collecting data from a large-scale sensor network (such as Smart Cities).

9.2. European Initiatives

9.2.1. FP7 & H2020 Projects

9.2.1.1. ERC-CLIM

Participants: Pierre David, Elian Dib, Christine Guillemot, Xin Su.

Light fields yield a rich description of the scene ideally suited for advanced image creation capabilities from a single capture, such as simulating a capture with a different focus and a different depth of field, simulating lenses with different apertures, for creating images with different artistic intents or for producing 3D views. Light fields technology holds great promises for a number of application sectors, such as photography, augmented reality, light field microscopy, but also surveillance, to name only a few.

The goal of the ERC-CLIM project is to develop algorithms for the entire static and video light fields processing chain, going from compact sparse and low rank representations and compression to restoration, high quality rendering and editing.

9.3. International Initiatives

9.3.1. Informal International Partners

- Reuben Farrugia, Prof. at the University of Malta has been one sabbatical year Sept. 2015-Aug. 2016) within the team, working on inverse problems (super-resolution, inpainting) for several applications.
- The study on guided image inpainting is carried out in collaboration with Prof. Pascal Frossard from EPFL (Ecole Polytechique Federal de Lausanne).

9.4. International Research Visitors

9.4.1. Visits of International Scientists

Sheila Hemami, Prof. at Northeastern University, Boston, has visited the team during three months (May 2016-July 2016), working on the problem of demultiplexing and decoding of micro-lenses based light fields.

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events Organisation

10.1.1.1. Member of Organizing Committees

- C. Guillemot has served as area chair for the European Signal Processing Conference (EUSIPCO), 2016.
- C. Guillemot co-organized (together with Joachim Weickert, Prof. Universität des Saarlandes, and Thomas Pock, Prof. TU Graz, and Gerlind Plonka-Hoch, Prof. Universität Göttingen) a Schloss Dagstuhl seminar on inpainting-based image compression (Nov. 13-18, 2016).

10.1.1.2. Chair of Conference Program Committees

- C. Guillemot was technical co-chair of the IEEE image, video and multi-dimensional signal processing workshop, 11-12 July 2016
- C. Guillemot was general co-chair of the IEEE International workshop on Multimedia Signal Processing (IEEE-MMSP), Montreal, 21-23 Sept. 2016.

10.1.1.3. Member of Conference Program Committees

- O. Le Meur has been a member of technical program committees of international conferences: EUVIP 2016, QoMEX 2016

- A. Roumy has been a member of the technical program committee of the ACCV 2016 workshop on New Trends in Image Restoration and Enhancement.

10.1.2. Journal

10.1.2.1. Member of the Editorial Boards

- C. Guillemot has been associate editor of the Eurasip International Journal on Image Communication (2010-2016).
- C. Guillemot is senior area editor of the IEEE Trans. on Image Processing.
- C. Guillemot is associate editor of the International Journal on Mathematical Imaging and Vision.
- O. Le Meur is member of the editorial board of the IET Image Processing Journal.
- T. Maugey has been Guest editor for the Special Issue on Interactive Multi-view Video Services: from acquisition to Rendering, IEEE Multimedia Communication Technical Committee letters, Vol. 11(2), March 2016 (Guest Editors: Erhan Ekmekcioglu, Thomas Maugey, Laura Toni)
- A. Roumy is associate editor of the Springer Annals of Telecommunications.

10.1.3. Invited Talks

- O. Le Meur has been invited for a talk at Harmonic (Rennes) on “Computational Modelling of visual attention”
- O. Le Meur participated to the TechnoConference at Inria. The presentation dealt with spatio-temporal video inpainting.
- T. Maugey has been invited for a talk at I3S (Nice, France) on “Enabling user to interactively select viewpoint : a challenge for 3D data compression ", May 2016.
- C. Guillemot gave an overview talk on image compression at the Dagstuhl seminar (13-18 Nov. 2016).

10.1.4. Leadership within the Scientific Community

- C. Guillemot is member of the IEEE IVMSP technical committee
- C. Guillemot is senior member of the steering committee of IEEE Trans. on Multimedia (2016-2018).

10.1.5. Scientific Expertise

- C. Guillemot is member as scientific expert of the CCRRDT (Regional Committee of Research and Technological Development) of the Brittany region.

10.1.6. Research Administration

- C. Guillemot has been (Sept. 2015-Dec. 2016) vice-chair of Inria’s evaluation committee.
- C. Guillemot is member of the “bureau du Comité des Projets”.
- T. Maugey has been member of the selection comitee for the assigment of the ministerial PhD grants at IRISA.
- A. Roumy served as a member of Board of Examiners (Comité de sélection) for an Associate Professor position (Maitre de Conférences) at CREATIS Polytech Lyon (MCU61-46 2016).
- A. Roumy served as a member of Board of Examiners (Comité de sélection) for an Associate Professor position (Maitre de Conférences) at ENSEIRB-MATMECA Bordeaux (MCF61-0127 2016).
- A. Roumy is a member of the Inria Joint Administrative Committee (CAP commission administrative paritaire).

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

Master: C. Guillemot, Image and video compression, 8 hours, M2 computer science, Univ. of Rennes 1, France.

Master: C. Guillemot, Image and video compression, 8 hours, M2 SISEA, Univ. of Rennes 1, France.

Master: O. Le Meur, Selective visual attention, 6 hours, M2, Univ. of Paris 8, France.

Master: O. Le Meur, Acquisition/Image Processing/Compression, 22 hours, M2 MITIC, Univ. of Rennes 1, France.

Engineer degree: O. Le Meur, Image Processing, video analysis and compression, 54 hours, ESIR2, Univ. of Rennes 1, France.

Engineer degree: O. Le Meur, Visual communication, 65 hours, ESIR3, Univ. of Rennes 1, France.

Professional training: O. Le Meur, Image Processing and OpenCV, 42 hours, Technicolor Rennes.

Engineering degree: A. Roumy, Compressive sensing, 17 hours, INSA Rennes, 5th year, Mathematical engineering, France.

Master: A. Roumy, Compressive sensing, 15 hours, ENSAI Bruz, Master Big Data, France.

10.2.2. Supervision of PhD defended during the year

PhD-VAE: P. Bordes, Adapting video compression to new formats, Univ. of Rennes 1, 18 Jan. 2016, C. Guillemot.

PhD : M. Le Pendu, Backward compatible approaches for the compression of high dynamic range videos, Univ. of Rennes 1, 17 March 2016, C. Guillemot (Cifre contract with Technicolor).

PhD : J. C. Ferreira, Algorithms for Super-resolution of Images based on Sparse Representation and Manifolds, co-tutelle Univ. of Rennes 1/ University of Uberlandia Brazil, 6 July 2016, C. Guillemot.

PhD: C. Chamaret, Color Harmony: experimental and computational modeling, University of Rennes 1, 28th of April, 2016.

10.2.3. Juries

- C. Guillemot has been member (rapporteur) of the PhD jury of:
  - T. Biatek, INSA-Rennes, Apr. 2016
  - A. Akl, Univ. of Bordeaux, Feb. 2016
- C. Guillemot has been member of the PhD jury of:
  - A. Gilles, INSA-Rennes, Sept. 2016
  - L. LE Magoarou, Univ. Rennes 1, Nov. 2016
- O. Le Meur has been member (rapporteur) of the jury of the PhD committee of:
  - Romain Cohendet, University Bretagne Loire, IRCCyN, 2016
  - Andrea Helo, University Paris Descartes, 2016
  - Ala Aboudib, Telecom Bretagne, 2016
  - Franck Chi, Telecom Bretagne, 2016

11. Bibliography

Publications of the year

Doctoral Dissertations and Habilitation Theses

[1] M. ALAIN.*A compact video representation format based on spatio-temporal linear embedding and epitome*, Université Rennes 1, January 2016, https://hal.inria.fr/tel-01261590.

[2] J. C. FERREIRA.*Algorithms for super-resolution of images based on sparse representation and manifolds*, Université Rennes 1, July 2016, https://tel.archives-ouvertes.fr/tel-01388977.

[3] M. LE PENDU.*Backward compatible approaches for the compression of high dynamic range videos*, Université Rennes 1, March 2016, https://tel.archives-ouvertes.fr/tel-01312901.

Articles in International Peer-Reviewed Journal

[4] P. BUYSSENS, O. LE MEUR, M. DAISY, D. TSCHUMPERLÉ, O. LÉZORAY.*Depth-guided disocclusion inpainting of synthesized RGB-D images*, in "IEEE Transactions on Image Processing", 2017, vol. 26, n^o 2, p. 525-538 [*DOI :* 10.1109/TIP.2016.2619263], https://hal.archives-ouvertes.fr/hal-01391065.

[5] J. C. FERREIRA, E. VURAL, C. GUILLEMOT.*Geometry-Aware Neighborhood Search for Learning Local Models for Image Superresolution*, in "IEEE Transactions on Image Processing", March 2016, vol. 25, n^o 3, 14, https://hal.inria.fr/hal-01388955.

[6] Y. GAO, G. CHEUNG, T. MAUGEY, P. FROSSARD, J. LIANG.*Encoder-Driven Inpainting Strategy in Multiview Video Compression*, in "IEEE Transactions on Image Processing", January 2016, vol. 25, n^o 1, p. 134-149, https://hal.inria.fr/hal-01217115.

[7] S. KHATTAK, T. MAUGEY, R. HAMZAOUI, S. AHMAD, P. FROSSARD.*Temporal and Inter-view Consistent Error Concealment Technique for Multiview plus Depth Video Broadcasting*, in "IEEE Transactions on Circuits and Systems for Video Technology", May 2016, vol. 26, n^o 5, p. 829-840 [*DOI :* 10.1109/TCSVT.2015.2418631], https://hal.inria.fr/hal-01137927.

[8] O. LE MEUR, A. COUTROT.*Introducing context-dependent and spatially-variant viewing biases in saccadic models*, in "Vision Research", 2016, vol. 121, p. 72 - 84 [*DOI :* 10.1016/J.VISRES.2016.01.005], https://hal.inria.fr/hal-01391745.

[9] M. LE PENDU, C. GUILLEMOT, D. THOREAU.*Inter-Layer Prediction of Color in High Dynamic Range Image Scalable Compression*, in "IEEE Transactions on Image Processing", August 2016, vol. 25, p. 3585 - 3596 [*DOI :* 10.1109/TIP.2016.2571559], https://hal.inria.fr/hal-01388961.

[10] T. MAUGEY, G. PETRAZZUOLI, P. FROSSARD, M. CAGNAZZO, B. PESQUET-POPESCU.*Reference view selection in DIBR-based multiview coding*, in "IEEE Transactions on Image Processing", April 2016, vol. 25, n^o 4, p. 1808-1819, https://hal.inria.fr/hal-01262258.

[11] E. VURAL, C. GUILLEMOT.*Out-of-sample generalizations for supervised manifold learning for classification*, in "IEEE Transactions on Image Processing", March 2016, vol. 25, n^o 3, 15, https://hal.inria.fr/hal-01388959.

Articles in Non Peer-Reviewed Journal

[12] A. ROUMY.*An Information theoretical problem in interactive Multi-View Video services*, in "IEEE Communications Society Multimedia Communications Technical Committee (ComSoc MMTC) E-Letter", March 2016, vol. 11, n^o 2, 6, https://hal.inria.fr/hal-01394864.

International Conferences with Proceedings

[13] M. ALAIN, C. GUILLEMOT, D. THOREAU, P. GUILLOTEL.*Learning Clustering-Based Linear Mappings for Quantization Noise Removal*, in "IEEE International Conference on Image Processing (ICIP) 2016", Phoenix, United States, September 2016 [*DOI :* 10.1109/ICIP.2016.7533151], https://hal.inria.fr/hal-01317625.

[14] T. DUMAS, A. ROUMY, C. GUILLEMOT.*Shallow sparse autoencoders versus sparse coding algorithms for image compression*, in "2016 IEEE International Conference on Multimedia and Expo (ICME 2016)", Seattle, WA, United States, July 2016, p. 1 - 6 [*DOI :* 10.1109/ICMEW.2016.7574708], https://hal.archives-ouvertes.fr/hal-01377907.

[15] R. FARRUGIA, C. GUILLEMOT.*Model and Dictionary guided Face Inpainting in the Wild*, in "ACCV workshop on New Trends in Image Restoration and Enhancement", Taipei, Taiwan, November 2016, 17, https://hal.inria.fr/hal-01388971.

[16] R. FARRUGIA, C. GUILLEMOT.*Robust Face Hallucination Using Quantization-Adaptive Dictionaries*, in "IEEE International Conference on Image Processing", Phoenix, United States, September 2016, 5, https://hal.inria.fr/hal-01388972.

[17] D. GOMMELET, A. ROUMY, C. GUILLEMOT, M. ROPERT, J. LETANOU.*Rate-distortion optimization of a tone mapping with SDR quality constraint for backward-compatible high dynamic range compression*, in "International Conference on Image Processing, ICIP", Phoenix, United States, 2016 IEEE International Conference on Image Processing (ICIP), September 2016, p. 1384-1388 [*DOI :* 10.1109/ICIP.2016.7532585], https://hal.inria.fr/hal-01377701.

[18] M. HOG, N. SABATER, C. GUILLEMOT.*Light Field Segmentation Using a Ray-Based Graph Structure*, in "European Conference on Computer Vision - ECCV", Amsterdam, Netherlands, October 2016, 16, https://hal.inria.fr/hal-01388965.

[19] O. LE MEUR, A. COUTROT.*How saccadic models help predict where we look during a visual task? Application to visual quality assessment*, in "SPIE Image Quality And System Performance", San Fransisco, United States, February 2016, https://hal.inria.fr/hal-01391750.

[20] M. RIZKALLAH, T. MAUGEY, C. YAACOUB, C. GUILLEMOT.*Impact of Light Field Compression on Refocused and Extended Focus Images*, in "EUSIPCO 2016 - 24th European Signal Processing Conference", Budapest, Hungary, August 2016 [*DOI :* 10.1109/EUSIPCO.2016.7760378], https://hal.inria.fr/hal-01377834.

[21] X. SU, T. MAUGEY, C. GUILLEMOT.*Graph-based representation for multiview images with complex camera configurations*, in "ICIP 2016 - IEEE International Conference on Image Processing", Phoenix, United States, September 2016, p. 1554 - 1558 [*DOI :* 10.1109/ICIP.2016.7532619], https://hal.inria.fr/hal-01378422.

National Conferences with Proceeding

[22] P. BUYSSENS, O. LE MEUR, M. DAISY, D. TSCHUMPERLÉ, O. LÉZORAY.*Désoccultation de cartes de profondeurs pour la synthèse de vues virtuelles*, in "RFIA 2016", Clermont-Ferrand, France, June 2016, https://hal.archives-ouvertes.fr/hal-01320967.

Scientific Books (or Scientific Book chapters)

[23] M. MANCAS, O. LE MEUR.*Applications of Saliency Models*, in "From Human Attention to Computational Attention. A Multidisciplinary Approach", M. MANCAS, V. P. FERRERA, N. RICHE, J. G. TAYLOR

(editors), Springer Series in Cognitive and Neural Systems, Springer, September 2016, vol. 10, p. 331-377 [*DOI :* 10.1007/978-1-4939-3435-5_18], https://hal.inria.fr/hal-01393254.

Other Publications

[24] A. COUTROT, O. LE MEUR.*Visual attention saccadic models: taking into account global scene context and temporal aspects of gaze behaviour*, August 2016, ECVP 2016 - European Conference on Visual Perception, Poster, https://hal.inria.fr/hal-01391751.

[25] M. HOG, N. SABATER, C. GUILLEMOT.*Super-rays for Efficient Light Field Processing*, December 2016, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01407852.

Project-Team SUMO

SUpervision of large MOdular and distributed systems

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

IN PARTNERSHIP WITH:
CNRS

Université Rennes 1

RESEARCH CENTER
Rennes - Bretagne-Atlantique

THEME
Proofs and Verification

Table of contents

Project-Team SUMO

Creation of the Team: 2013 January 01, updated into Project-Team: 2015 January 01

Keywords:

Computer Science and Digital Science:

1.2.2. - Supervision
1.3. - Distributed Systems
2.3.2. - Cyber-physical systems
2.4.2. - Model-checking
4.5. - Formal methods for security
6.4. - Automatic control
7.1. - Parallel and distributed algorithms
7.3. - Optimization
7.4. - Logic in Computer Science
7.8. - Information theory
7.14. - Game Theory

Other Research Topics and Application Domains:

1.1.3. - Cellular biology
1.1.9. - Bioinformatics
5.2.2. - Railway
6.2. - Network technologies
6.3.3. - Network Management

1. Members

Research Scientists

Éric Fabre [Team leader, Researcher, Inria, HDR]
Éric Badouel [Inria, Researcher, HDR]
Nathalie Bertrand [Inria, Researcher, HDR]
Blaise Genest [CNRS, Researcher, HDR]
Loïc Hélouët [Inria, Researcher, HDR]
Thierry Jéron [Inria, Researcher, HDR]
Hervé Marchand [Inria, Researcher]
Nicolas Markey [CNRS, Researcher, HDR]
Ocan Sankur [CNRS, Researcher]

Technical Staff

Olivier Bache [Inria, from April to September 2016]

PhD Students

Hugo Bazille [Inria]
Sihem Cherrared [Orange Labs, from Sep 2016, granted by CIFRE]
Paulin Fournier [Univ. Rennes I, until Aug 2016]
Abd El Karim Kecir [Alstom, granted by CIFRE]
Engel Lefaucheux [Univ. Rennes I]
Matthieu Pichené [Inria]

Post-Doctoral Fellow
Sucheendra Palaniappan [Inria, until Sep 2016]

Visiting Scientists
Robert Nsaibirni [University of Yaoundé I, from September 2016]
Shauna Laurene Ricker [Prof. Mount Allison Univ., Canada, until Jul 2016]

Administrative Assistant
Laurence Dinh [Inria]

Others
Vincent Aubry [ENS Paris, from Jun 2016 until Jul 2016]
Christophe Morvan [Ass. Prof., Univ. Paris Est Marne La Vallée]
Arthur Queffelec [Inria, from May 2016 until Jul 2016]
Jérémy Thibault [Inria, from May 2016 until Jul 2016]

2. Overall Objectives

2.1. Overall objectives

Most software driven systems we commonly use in our daily life are huge hierarchical assemblings of components. This observation runs from the micro-scale (multi-core chips) to the macro-scale (data centers), and from hardware systems (telecommunication networks) to software systems (choreographies of web services). The main characteristics of these pervasive applications are size, complexity, heterogeneity, and modularity (or concurrency). Besides, several such systems are actively used before they are fully mastered, or they have grown so much that they now raise new problems that are hardly manageable by human operators. While these systems and applications are becoming more essential, or even critical, the need for their *reliability, efficiency* and *manageability* becomes a central concern in computer science. The main objective of SUMO is to develop theoretical tools to address such challenges, according to the following axes.

2.1.1. Necessity of quantitative models.

Several disciplines in computer science have of course addressed some of the issues raised by large systems. For example, formal methods (essentially for verification purposes), discrete event systems (diagnosis, control, planning, and their distributed versions), but also concurrency theory (modelling and analysis of large concurrent systems). Practical needs have oriented these methods towards the introduction of quantitative aspects, such as time, probabilities, costs, and their combinations. This approach drastically changes the nature of questions that are raised. For example, verification questions become the reachability of a state in a limited time, the average sojourn duration in a state, the probability that a run of the system satisfies some property, the existence of control strategies with a given winning probability, etc. In this setting, exact computations are not always appropriate as they may end up with unaffordable complexities, or even with undecidability. Approximation strategies then offer a promising way around, and are certainly also a key to handling large systems. Discrete event systems approaches follow the same trend towards quantitative models. For diagnosis aspects, one is interested in the most likely explanations to observed malfunctions, in the identification of the most informative tests to perform, or in the optimal placement of sensors. For control problems, one is of course interested in optimal control, in minimizing communications, in the robustness of the proposed controllers, in the online optimization of QoS (Quality of Service) indicators, etc.

2.1.2. Specificities of distributed systems.

While the above questions have already received partial answers, they remain largely unexplored in a distributed setting. We focus on structured systems, typically a network of dynamic systems with known interaction topology, the latter being either static or dynamic. Interactions can be synchronous or asynchronous. The state space explosion raised by such systems has been addressed through two techniques. The first one consists in adopting true concurrency models, which take advantage of the parallelism to reduce the size of the trajectory sets. The second one looks for modular or distributed “supervision" methods, taking the shape of a network of local supervisors, one per component. While these approaches are relatively well understood, their mixing with quantitative models remains a challenge (as an example, there exists no proper setting assembling concurrency theory with stochastic systems). This field is largely open both for modeling, analysis and verification purposes, and for distributed supervision techniques. The difficulties combine with the emergence of data driven distributed systems (as web services or data centric systems), where the data exchanged by the various components influence both the behaviors of these components and the quantitative aspects of their reactions (e.g. QoS). Such systems call for symbolic or parametric approaches for which a theory is still missing.

2.1.3. New issues raised by large systems.

Some existing distributed systems like telecommunication networks, data centers, or large scale web applications have reached sizes and complexities that reveal new management problems. One can no longer assume that the model of the managed systems is static and fully known at any time and any scale. To scale up the management methods to such applications, one needs to be able to design reliable abstractions of parts of the systems, or to build dynamically a part of their model, following the needs of the management functions to realize. Besides, one does not wish to define management objectives at the scale of each single component, but rather to pilot these systems through high-level policies (maximizing throughput, minimizing energy consumption, etc.). These distributed systems and management problems have connections with other approaches for the management of large structured stochastic systems, such as Bayesian networks (BN) and their variants. The similarity can actually be made more formal: inference techniques for BN rely on the concept of conditional independence, which has a counterpart for networks of *dynamic* systems and is at the core of techniques like distributed diagnosis, distributed optimal planning, or the synthesis of distributed controllers. The potential of this connection is largely unexplored, but it suggests that one could derive from it good approximate management methods for large distributed dynamic systems.

3. Research Program

3.1. Analysis and verification of quantitative systems

The overall objective of this axis is to develop the quantitative aspects of formal methods while maintaining the tractability of verification objectives and progressing toward the management of large systems. This covers the development of relevant modeling formalims, to nicely weave time, costs and probabilities with existing models for concurrency. We plan to further study time(d) Petri nets, networks of timed automata (with synchronous or asynchronous communications), stochastic automata, partially observed Markov decision processes, etc. A second objective is to develop verification methods for such quantitative systems. This covers several aspects: quantitative verification questions (compute an optimal scheduling policy), boolean questions on quantitative features (deciding whether some probability is greater than a threshold), robustness issues (will a system have the same behaviors if some parameter is slightly altered), etc. Our goal is to explore the frontier between decidable and undecidable problems, or more pragmatically tractable and untractable problems. Of course, there is a tradeoff between the expressivity and the tractability of a model. Models that incorporate distributed aspects, probabilities, time, etc, are typically untractable. In such a case, abstraction or approximation techniques are a work around that we will explore.

Here are some more detailed topics that we place in our agenda

- analysis of diagnosability and opacity properties for stochastic systems
- verification of time(d) Petri nets
- robustness analysis for timed or/and stochastic systems
- abstraction techniques for quantitative systems

3.2. Control of quantitative systems

The main objective of this research axis is to explore the quantitative and/or distributed extensions of classical control problems. We envision control in its widest meaning of driving a system in order to guarantee or enforce some extra property (i.e. not guaranteed by the system alone), in a partially or totally observed setting. This property can either be logical (e.g. reachability or safety) or quantitative (e.g. reach some performance level). These problems have of course an offline facet (e.g. controller design, existence of a policy/strategy) and an online facet (e.g. algorithm to select some optimal action at runtime).

Our objectives comprise classical controler synthesis for discrete event systems, with extensions to temporal/stochastic/reward settings. They also cover maintaining or maximizing extra properties as diagnosability or opacity, for example in stochastic systems. We also target further analysis of POMDPs (partially observed Markov decision processes), and multi-agent versions of policy synthesis relying on tools from game theory. We aim at adressing some control problems motivated by industrial applications, that raise issues like the optimal control of timed and stochastic discrete event systems, with concerns like robustness to perturbations and multicriteria optimization. Finally, we also plan to work on modular testing, and on runtime enforcement techniques, in order to garantee extra logical and temporal properties to event flows.

3.3. Management of large or distributed systems

The generic terms of “supervision” or “management” of distributed systems cover problems like control, diagnosis, sensor placement, planning, optimization, (state) estimation, parameter identification, testing, etc. This research axis examines how classical settings for such problems can scale up to large or distributed systems. Our work will be driven by considerations like : how to take advantage of modularity, how to design approximate management algorithms, how to design relevant abstractions to make large systems more tractable, how to deal with models of unknown size, how to design mechanisms to obtain relevant models, etc.

As more specific objectives, let us mention:

- Parametric systems. How to verify properties of distributed systems with an unknown number of components.
- Approximate management methods. We will explore the extension of ideas developed for Bayesian inference in large scale stochastic systems (such as turbo-algorithms for example) to the field of modular dynamic systems. When component interactions are sparse, even if exact management methods are unaccessible (for diagnosis, planning, control, etc.), good approximations based on local computations may be accessible.
- Model abstraction. We will explore techniques to design more tractable abstractions of stochastic dynamic systems defined on large sets of variables.
- Self-modeling, which consists in managing large scale systems that are known by their building rules, but which specific managed instance is only discovered at runtime, and on the fly. The model of the managed system is built on-line, following the needs of the management algorithms.
- Distributed control. We will tackle issues related to asynchronous communications between local controllers, and to abstraction techniques allowing to address large systems.
- Test and enforcement. We will tackle coverage issues for the test of large systems, and the test and enforcement of properties for timed models, or for systems handling data.

3.4. Data driven systems

Data-driven systems are systems whose behavior depends both on explicit workflows (scheduling and durations of tasks, calls to possibly distant services,...) and on the data processed by the system (stored data, parameters of a request, results of a request,...). This family of systems covers workflows that convey data (business processes or information systems), transactional systems (web stores), large databases managed with rules (banking systems), collaborative environments (crowds, health systems), etc. These systems are distributed, modular, and open: they integrate components and sub-services distributed over the web and accept requests from clients. Our objective is to provide validation and supervision tools for such systems. To achieve this goal, we have to solve several challenging tasks:

- provide realistic models, and sound automated abstraction techniques, to reason on models that are reasonable abstractions of real systems. These models should be able to encompass modularity, distribution, in a context where workflows and data aspects are tightly connected.
- address design of data driven systems in a declarative way: declarative models are another way to handle data-driven systems. Rather than defining the explicit workflows and their effects on data, rule-based models state how actions are enacted in terms of the shape (pattern matching) or value of the current data. We think that distributed rewriting rules or attributed grammars can provide a practical yet formal framework for maintenance, by providing a solution to update mandatory documentation during the lifetime of an artifact.
- provide tractable solutions for validation of models. Frequent issues are safety questions (can a system reach some bad configuration?), but also liveness (workflows progess), ... These questions should not only remain decidable on our models, but also with efficient computational methods.
- address QoS management in large reconfigurable systems: Data driven distributed systems often have constraints in terms of QoS. This QoS questions adresse performance issues, but also data quality. This calls for an analysis of quantitative features and for reconfiguration techniques to meet desired QoS.

4. Application Domains

4.1. Smart transportation systems

The smart cities trend aims at optimizing all functions of future cities with the help of digital technologies. We focus on the segment of urban trains, which will evolve from static and scheduled offers to reactive and eventually on demand transportation offers. We address two challenges in this field. The first one concerns the optimal design of robust subway lines. The idea is to be able to evaluate, at design time, the performance of time tables and of different regulations policies. In particular, we focus on robustness issues: how small perturbations and incidents can be accomodated by the system, and how fast return to normality occurs, when does the system become unstable. The second challenge concerns the design of new robust regulation strategies to optimize delays, recovery times, and energy consumption at the scale of a full subway line. These problems involve large scale discrete event systems, with temporal and stochastic features, and translate into robustness assessment, stability analysis and joint numerical/combinatorial optimization problems on the trajectories of these systems.

4.2. Management of telecommunication networks and of data centers

Telecommunication network management is a rich provider of research topics for the team, and some members of Sumo have a long background of contacts and transfer with industry in this domain. Networks are typical examples of large distributed dynamic systems, and their management raises numerous problems ranging from diagnosis (or root cause analysis), to optimization, reconfiguration, provisioning, planning, verification, etc. They also bring new challenges to the community. For example on the modeling side, building or learning a network model is a complex task, specifically because these models should reflect features like the layering, the multi-resolution view of components, the description of both functions, protocols and configuration, and they should reflect as well dynamically changing architectures. Besides modeling, management algorithms are also challenged by features like the size of systems, the need to work on abstractions, on partially known models, on open (multi-tenant) systems, on dynamically changing systems, etc. The networking technology is now evolving toward software defined networks, virtualized network functions, which reinforces the need for more automation in the management of such systems.

Data centers are another example of large scale modular dynamic and reconfigurable systems: they are composed of thousands of servers, on which virtual machines are activated, migrated, resized, etc. Their management covers issues like trouble shooting, reconfiguration, optimal control, in a setting where failures are frequent and mitigated by the performance of the management plane. We have a solid background in the coordination of the various autonomic managers that supervise the different functions/layers of such systems (hardware, middleware, web services,...) . Virtualization technologies now reach the domain of networking, and telecommunication operators/vendors evolve towards providers of distributed open clouds. This convergence of IT and networking strongly calls for new management paradigms, which is an opportunity for the team.

This application domain will be revived in the team by a collaboration with Orange Labs (1 CIFRE PhD in the common lab Orange/Inria) and a collaboration with Nokia Bell Labs (1 CIFRE PhD, and participation to the joint research team “Softwarization of Everything” of the common lab Nokia Bell Labs/Inria).

4.3. Collaborative workflows

A current trend is to involve end-users in collection and analysis of data. Exemples of this trend are contributive science, crisis management systems, and crowds. All these applications are data-centric and user-driven. They are often distributed and involve complex and sometimes dynamic workflows. In many cases, there are strong interactions between data and control flows: indeed, decisons taken to decide of the next tasks to be launched highly depend on collected data. For instance, in an epidemic surveillance system, the aggregation of various reported disease cases may trigger alerts. Another example is crowds where user skills are used to complete tasks that are better performed by humans than computers. In return, this needs to address imprecise and sometimes unreliable answers. We address several issues related to complex workflows and data. We study declarative and dynamic models that can handle workflows, data, uncertainty, and competences management. Once these models are mature enough, we plan to experiment them on real use cases from contributive science, health management systems, and crowd platforms using prototypes. We also plan to define abstaction schemes allowing formal reasonning on these systems.

4.4. Systems Biology

A quite new topic in SUMO is about Systems Biology. In systems biology, many continuous variables interact together. Biological systems are thus good representatives for large complex quantitative systems, for which we are developing analysis and management methods. For instance, the biological pathway of apoptosis explain how many molecules interact inside a cell, triggered by some outside signal (drug, etc.), eventually leading to the death of the cell through apoptosis. While intrinsically quantitative in nature, data are usually noisy and problems need not be answered with ultimate precision. It thus seems reasonable to resort to approximations in order to handle the state space explosion resulting from the high dimensionality of biological systems.

We are developing models and abstraction tools for system biology. Studying these models suggests new reduction methods, such as considering populations instead of explicitly representing every single element into play (be it cells, molecules, etc): we thus develop algorithm handling population symbolically, either in a continuous (distributions) or a discrete (parametric) way. An intermediate goal is to speed-up analysis of such systems using abstractions, and a long term goal is to develop top down model-checking methods that can be run on these abstractions.

5. Highlights of the Year

5.1. Highlights of the Year

Start-up creation. Christophe Morvan (Ass. Prof. Univ. Paris Est Marne la Vallée) has been hosted by Sumo for several years for his research activities. In 2016, he created Open Agora with two other computer scientists.

The company develops a software suite to help the decision process in large structures. It offers tools to structure discussions, voting mechanisms, and automated argument summaries. The company will maintain connections with the team for the development of GAGs (Guarded Attributed Grammars) that are instrumental in the automated summary tools.

New team member. Nicolas Markey (DR CNRS) recently joined the team, after several years in LSV (*Laboratoire Spécification et Vérification*), Cachan. Nicolas will reinforce the activities of the team in the modeling and analysis of timed systems, abstraction techniques and game theory.

6. New Software and Platforms

6.1. Active Workspaces

KEYWORDS: Guarded attribute grammar - Active workspace - Artifact centric workflow system

SCIENTIFIC DESCRIPTION

Tool for computer supported cooperative work where a user's workspace is given by an active structured repository containing the pending tasks together with information needed to perform the tasks. Communication between active workspaces is asynchronous using message passing. The tool is based on the model of guarded attribute grammars [44]. Late in 2015 Éric Badouel produced in Haskell a software prototype implementing active workspaces based on Guarded Attribute Grammars (GAGs).

Concurrently, Christophe Morvan was beginning a startup project consisting in making on-line collective decision making tools: *Open Agora*. This project included collaboration workspaces for people participating in constructing possible decisions. There was a natural connection between the prototype, and the startup project.

In order to make industrial use of the GAG prototype, Olivier Bache (already involved in the Open agora project) applied to a 6 month InriaHub program (between April and September 2016). During these 6 months he bundeled the prototype into an API (also programmed in Haskell) and developped a web infrastructure, based on the PHP framework, to allow the interaction with Active Workspaces in a browser. This developpement will be licenced to Open Agora SAS after its creation expected in January 2017.

FUNCTIONAL DESCRIPTION

Prototype in Haskell of user's active workspaces based on Guarded Attribute Grammars.

- Author: Eric Badouel
- Contact: Eric Badouel
- URL: http://people.rennes.inria.fr/Eric.Badouel/Research/ActiveWorkspaces.html

6.2. SIMSTORS

SIMSTORS is a simulator for regulated stochastic timed Petri nets. These Petri nets are a variant of stochastic and timed nets, whose execution is controlled by a regulation policy an a predetermined theoretical schedule. The role of the regulation policy is to control the system to realize the schedule with the best possible precision. This software allows not only for step by step simulation, but also for performance analysis of systems such as production cells or train systems.

SIMSTORS was used successfully during a collaboration with Alstom transport to model existing urban railway systems and their regulation schemes. Alstom transport is willing to transfer this software and use it during early design phase of regulation algorithms in their metro lines. This year, the software has been extended to consider headway management.

- Participants: Loïc Hélouët and Karim Kecir
- Contact: Loïc Hélouët
- URL: http://www.irisa.fr/sumo/Software/SIMSTORS/

7. New Results

7.1. Analysis and verification of quantitative systems

7.1.1. Quantitative verification of distributions of stochastic models

Participant: Blaise Genest.

In [24], we obtained conditions under which quantitative verification of distributions of stochastic systems is decidable. This is a challenging question as for general Markov Chains, verification of distribution is Skolem-complete, a problem on linear recurrence sequences whose decidability is a long-standing problem open for 40 years. In this paper, we approach this problem by studying the languages generated by Markov Chains, whose regularity would entail the decidability of quantitative verification. Given an initial distribution, we represent the trajectory of Markov Chain over time as an infinite word over a finite alphabet, where the n^{th} letter represents a probability range after n steps. We extend this to a language of trajectories (a set of words), one trajectory for each initial distribution from a (possibly infinite) set. We show that if the eigenvalues of the transition matrix associated with the Markov Chain are all distinct positive real numbers, then the language is *effectively regular*. Further, we show that this result is at the boundary of regularity, as non-regular languages can be generated when the restrictions are even slightly relaxed. The regular representation of the language allows us to reason about more general properties, e.g., robustness of a regular property in a neighbourhood around a given distribution.

7.1.2. Diagnosability of repairable faults

Participants: Éric Fabre, Loïc Hélouët, Hervé Marchand, Engel Lefaucheux.

For (partially observable) discrete event systems, diagnosability characterizes the ability to detect the occurrence of a permanent fault in bounded time after it occurs, given the observations available on that system. Diagnosability can be decided in polynomial time, relying on the so-called twin-machine construction. We have examined the case of repairable faults, and a notion of diagnosability that requires the detection of the fault before it is repaired. It was proved in [35]that diagnosability is a PSpace complete problem.

7.1.3. Diagnosability of stochastic systems

Participants: Éric Fabre, Blaise Genest, Hugo Bazille, Ocan Sankur.

Diagnosis of partially observable stochastic systems prone to faults was introduced in the late nineties. Diagnosability, i.e. the existence of a diagnoser, may be specified in different ways: (1) exact diagnosability (called A-diagnosability) requires that almost surely a fault is detected and that no fault is erroneously claimed while (2) approximate diagnosability (called ε-diagnosability) allows a small probability of error when claiming a fault and (3) accurate approximate diagnosability (called AA-diagnosability) requires that this error threshold may be chosen arbitrarily small. In a recent work [27], we focused on approximate diagnoses. We first refined the almost sure requirement about finite delay introducing a uniform version and showing that while it does not discriminate between the two versions of exact diagnosability this is no more the case in approximate diagnosis. We then gave a complete picture of relations between the different diagnosability specifications for probabilistic systems and establish characterisations for most of them in the finite-state case. Based on these characterisations, we developped decision procedures, studied their complexity and proved their optimality. We also designed synthesis algorithms to construct diagnosers and we analysed their memory requirements. Finally we established undecidability of the diagnosability problems for which we provided no characterisation. Notably, we proved the AA-diagnosability problem to be undecidable, answering a longstanding open question.

In another work [28], we investigated semantical and computational issues for exact notions of diagnosability in the context of infinite-state probabilistic systems. We first showed established a characterisation of the so-called FF-diagnosability using a $G\delta$ set (instead of an open set for finite-state systems) and also for two other notions, IF- and IA-diagnosability, when models are finitely branching. We also proved that surprisingly the last notion, FA-diagnosability, cannot be characterised in this way even in the finitely branching case. Then we applied our characterisations for a partially observable probabilistic extension of visibly pushdown automata, yielding EXPSPACE procedures for solving diagnosability problems. In addition, we establish some computational lower bounds and show that slight extensions of these probabilistic visibly pushdown automata lead to undecidability.

7.1.4. Analysing decisive stochastic processes

Participant: Nathalie Bertrand.

In 2007, Abdulla et al. introduced the elegant concept of decisive Markov chain. Intuitively, decisiveness allows one to lift the good properties of finite Markov chains to infinite Markov chains. For instance, the approximate quantitative reachability problem can be solved for decisive Markov chains (enjoying reasonable effectiveness assumptions) including probabilistic lossy channel systems and probabilistic vector addition systems with states. In a recent work [26], we extended the concept of decisiveness to more general stochastic processes. This extension is non trivial as we consider stochastic processes with a potentially continuous set of states and uncountable branching (common features of real-time stochastic processes). This allowed us to obtain decidability results for both qualitative and quantitative verification problems on some classes of real-time stochastic processes, including generalized semi-Markov processes and stochastic timed automata.

7.1.5. Concurrent timed systems

Participants: Loïc Hélouët, Blaise Genest.

Adding real time information to Petri net models often leads to undecidability of classical verification problems such as reachability and boundedness. For instance, models such as Timed-Transition Petri nets (TPNs) [47] are intractable except in a bounded setting. On the other hand, the model of Timed-Arc Petri nets [50] enjoys decidability results for boundedness and control-state reachability problems at the cost of disallowing urgency (the ability to enforce actions within a time delay).

We have addressed semantics variants of time and timed Petri nets to obtain concurrent models with interesting expressive power, but yet allowing decidability of verification and robustness questions. Robustness of timed systems aims at studying whether infinitesimal perturbations in clock values can result in new discrete behaviors. A model is robust if the set of discrete behaviors is preserved under arbitrarily small (but positive) perturbations.

In [25] we have considered time in Petri nets under a strong semantics with multiple enabling of transitions. We focus on a structural subclass of unbounded TPNs, where the underlying untimed net is free-choice, and show that it enjoys nice properties under a multi-server semantics. In particular, we showed that the questions of fireability (whether a chosen transition can fire), and termination (whether the net has a non-terminating run) are decidable for this class. We then consider the problem of robustness under guard enlargement [48], i.e., whether a given property is preserved even if the system is implemented on an architecture with imprecise time measurement. Unlike in [15], where decidability of several problems is obtained for bounded classes of nets, we showed that robustness of fireability is decidable for unbounded free choice TPNs with a multi-server semantics.

The robustness of time Petri nets was addressed in [15] by considering the model of parametric guard enlargement which allows time-intervals constraining the firing of transitions in TPNs to be enlarged by a (positive) parameter. We show that TPNs are not robust in general and checking if they are robust with respect to standard properties (such as boundedness, safety) is undecidable. We then extend the marking class timed automaton construction for TPNs to a parametric setting, and prove that it is compatible with guard enlargements. We apply this result to the (undecidable) class of TPNs which are robustly bounded (i.e., whose finite set of reachable markings remains finite under infinitesimal perturbations): we provide two decidable robustly bounded subclasses, and show that one can effectively build a timed automaton which is timed bisimilar even in presence of perturbations. This allows us to apply existing results for timed automata to these TPNs and show further robustness properties.

The goal of [23] is to investigate decidable classes of Petri nets with time that capture some urgency and still allow unbounded behaviors, which go beyond finite state systems. We have shown, up to our knowledge, the first decidability results on reachability and boundedness for Petri net variants that combine unbounded places, time, and urgency. For this, we have introduced the class of Timed-Arc Petri nets with restricted Urgency, where urgency can be used only on transitions consuming tokens from bounded places. We showed that control-state reachability and boundedness are decidable for this new class, by extending results from Timed-Arc Petri nets (without urgency) [43]. Our main result concerns (marking) reachability, which is undecidable for both TPNs (because of unrestricted urgency) [46] and Timed-Arc Petri Nets (because of infinite number of "clocks") [49]. We obtained decidability of reachability for unbounded TPNs with restricted urgency under a new, yet natural, timed-arc semantics presenting them as Timed-Arc Petri Nets with restricted urgency. Decidability of reachability under the intermediate marking semantics is also obtained for a restricted subclass.

7.1.6. Petri nets realizability

Participants: Loïc Hélouët, Abd El Karim Kecir.

We considered in [30] the realizability of urban train schedules by stochastic concurrent timed systems. Schedules are high level views of desired timetables that a metro system should implement. They are represented as partial orders decorated with timing constraints. Train systems are represented as elementary stochastic time Petri nets. We have first considered logical realizability: a schedule is realizable by a net $\mathcal{N}$ if it embeds in a time process of $\mathcal{N}$ that satisfies all its constraints. However, with continuous time domains, the probability of a time process that realizes a schedule is null. We have extended the former notion of realizability to consider probabilistic realizability of schedules up to some imprecision α. This probabilistic realizability holds if the probability that $\mathcal{N}$ logically realizes S with constraints enlarged by α time units is strictly positive. We have shown that upon a sensible restriction guaranteeing time progress (systems can not perform an arbitrary number of actions within a single time unit), logical and probabilistic realizability of a schedule can be checked on the finite set of symbolic prefixes extracted from a bounded unfolding of the net. We have provided a construction technique for these prefixes and shown that they represent all time processes of a net occurring up to a given maximal date. We have then shown how to verify existence of an embedding and compute the probability of its realization.

7.2. Control of quantitative systems

7.2.1. Smart regulation for urban trains

Participants: Éric Fabre, Loïc Hélouët, Hervé Marchand, Abd El Karim Kecir.

The regulation of subway lines consists in accomodating small random perturbations in transit times as well as more impacting incidents, by playing on continuous commands (transit times and dwell times) and by making more complex decisions (insertions or extractions of trains, changes of missions, overpassing, shorter returns, etc.). The objectives are multiple : ensuring the regularity and punctuality of trains, adapting to transportation demand, minimizing energy consumption, etc. We have developed an event-based control strategy that aims at equalizing headways on a line. This distributed control strategy is remarquably robust to perturbations and reactive enough to accomodate train insertions/extractions. We have also developed another approach based on event graphs in order to optimally interleave trains at a junction.

7.2.2. Games and reactive synthesis

Participant: Ocan Sankur.

In game theory, a strategy is *dominated* by another one if the latter systematically yields a payoff as good as the former, while also yielding a better payoff in some cases. A strategy is *admissible* if it is not dominated. This notion is well studied in game theory and is useful to describe the set of strategies that are "reasonable" whose choice can be justified. Recent works studied this notion in graph games with omega-regular objectives and investigated its applications in controller synthesis. For multi-agent controller synthesis, admissibility can be used as a hypothesis on the behaviors of each agent, thus enabling a compositional reasoning framework for controller synthesis. In [29], we investigate this framework for quantitative graph games. We characterize admissible strategies, study their existence, and give an effective characterization of the set of paths that are compatible with admissible payoffs. This is then used to derive algorithms for model checking under admissibility, but also assume-admissible synthesis.

In [21], we present the reactive synthesis competition (SYNTCOMP), a long-term effort intended to stimulate and guide advances in the design and application of synthesis procedures for reactive systems. The first iteration of SYNTCOMP is based on the controller synthesis problem for finite-state systems and safety specifications. We provide an overview of this problem and existing approaches to solve it, and report on the design and results of the first SYNTCOMP. This includes the definition of the benchmark format, the collection of benchmarks, the rules of the competition, and the five synthesis tools that participated. We present and analyze the results of the competition and draw conclusions on the state of the art. Finally, we give an outlook on future directions of SYNTCOMP.

In the invited [22], we summarize new solution concepts useful for the synthesis of reactive systems that we have introduced in several recent publications. These solution concepts are developed in the context of non-zero sum games played on graphs. They include the assume-admissible synthesis on Boolean games, synthesis under multiple environments for Markov decision processes, and multi-objective synthesis with probability thresholds for Markov decision processes with multi-dimensional weights. They are part of the contributions obtained in the inVEST project funded by the European Research Council.

7.2.3. Runtime enforcement

Participants: Hervé Marchand, Thierry Jéron.

In the [20] we generalize our line of work on runtime enforcement for timed properties. Runtime enforcement is a verification/validation technique aiming at correcting possibly incorrect executions of a system of interest. In this work we consider enforcement monitoring for systems where the physical time elapsing between actions matters. Executions are thus modelled as timed words (i.e., sequences of actions with dates). We consider runtime enforcement for timed specifications modelled as timed automata. Our enforcement mechanisms have the power of both delaying events to match timing constraints, and suppressing events when no delaying is appropriate, thus possibly allowing for longer executions. To ease their design and their correctness-proof, enforcement mechanisms are described at several levels: enforcement functions that specify the input-output behaviour in terms of transformations of timed words, constraints that should be satisfied by such functions, enforcement monitors that describe the operational behaviour of enforcement functions, and enforcement algorithms that describe the implementation of enforcement monitors.

This year we went one step ahead [33] and consider predictive runtime enforcement, where the system is not entirely black-box, but we know something about its behavior. This *a priori* knowledge about the system allows to output some events immediately, instead of delaying them until more events are observed, or even blocking them permanently. This in turn results in better enforcement policies. We also show that if we have no knowledge about the system, then the proposed enforcement mechanism reduces to a classical non-predictive runtime enforcement framework. All our results are formalized and proved in the Isabelle theorem prover.

7.2.4. Decentralized control

Participant: Hervé Marchand.

In collaboration with Laurie Ricker, we have been interested in decentralized control of discrete event systems. In decentralized discrete-event system (DES) architectures, agents fuse their local decisions to arrive at the global decision. The contribution of each agent to the final decision is never assessed; however, it may be the case that only a subset of agents, i.e., a (static) coalition, perpetually contribute towards the correct final decisions. In casting the decentralized DES control (with and without communication) problem as a cooperative game, it is possible to quantify the average contribution that each agent makes towards synthesizing the overall correct control strategy. Specifically, we explore allocations that assess contributions of non-communicating and communicating controllers for this class of problems. This allows a quantification of the contribution that each agent makes to the coalition with respect to decisions made solely based on its partial observations and decisions made based on messages sent to another agent(s) to facilitate a correct control decision [34].

7.3. Management of large distributed systems

7.3.1. Non-interference in partial order models

Participant: Loïc Hélouët.

We obtained new results on security issues such as non-interference [41]. Noninterference (NI) is a property of systems stating that confidential actions should not cause effects observable by unauthorized users. Several variants of NI have been studied for many types of models but rarely for true concurrency or unbounded models. In [45], we had already demonstrated the discriminating power of partial orders, and investigated NI for High-level Message Sequence Charts (HMSCs), a partial order language for the description of distributed systems. We had proposed a general definition of security properties in terms of equivalence among observations of behaviors, and showed that equivalence, inclusion, and NI properties were undecidable for HMSCs. We defined a new formalism called *partial order automata*, that captures natural observations of distributed systems, and in particular observations of HMSCs. It generalizes HMSCs and permits assembling partial orders. We have then considered subclasses of partial order automata and HMSCs for which Non-Interference is decidable. This allowed us to exhibit more classes of HMSCs for which NI is decidable. Finally, we have defined weaker local Non-interfernce properties, describing situations where a system is attacked by a single agent, and shown that local NI is decidable. We have then refined local NI to a finer notion of causal NI that emphasizes causal dependencies between confidential actions and observations and extended it to causal NI with (selective) declassification of confidential events, which allows to consider that confidential actions need can be kept secret for a limited duration and can then be declassified. Checking whether a system satisfies local and causal NI and their declassified variants are PSPACE-complete problems.

7.3.2. Simulations for stochastic abstractions of large systems

Participants: Éric Fabre, Blaise Genest, Matthieu Pichené.

In [32], we developped a new simulation strategy to accurately simulate DBNs (Dynamic Bayesian Networks) obtained as stochastic abstractions of large systems. The DBN abstractions are given under the form of probability tables, describing the probability for a variable to take a given value given the values of some variables at the previous time point. To be able to handle large systems with many variables, there is a table for each variable (coupling between variable is not explicitely represented). This creates discrepencies when simulating variables independently. Our new algorithm simulates tuples of variables together by looking ahead for such discrepencies in order to avoid them. Such simulations are still efficient, and match more faithfully the original systems.

7.4. Data driven systems

7.4.1. Structured data nets

Participants: Éric Badouel, Loïc Hélouët, Christophe Morvan.

In [16] we proposed a Petri net extension, called Structured Data Nets (SDN), that describes transactional systems with data. In these nets, tokens are semi-structured documents. Each transition is attached to a query, guarded by patterns, (logical assertions on the contents of its preset) and transforms tokens.

We define SDNs and their semantics and consider their formal properties: coverability of a marking, termination and soundness of transactions.

Unrestricted SDNs are Turing complete, so these properties are undecidable. We thus used an order on documents, and showed that under reasonable restrictions on documents and on the expressiveness of patterns and queries, SDNs are well-structured transition systems, for which coverability, termination and soundness are decidable.

7.4.2. An active workspace model for disease surveillance

Participant: Éric Badouel.

Flexibility and change at both design- and run-time are fast becoming the Rule rather than the Exception in Business Process Models. This is attributed to the continuous advances in domain knowledge, the increase in expert knowledge, and the diverse and heterogeneous nature of contextual variables. Such processes are characterized by collaborative work and decision making between users with heterogeneous profiles on a processes designed on-the-fly. A model for such processes should thus natively support human interactions. We showed in [31] how the Active Workspaces model proposed [44] for distributed collaborative systems supports these interactions.

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Contracts with Industry

Joint Alstom-Inria research lab: Several researchers of SUMO are involved in the joint research lab of Alstom and Inria, in a common research team called P22. On Alstom side, this joint research team involves researchers of the ATS division (Automatic Train Supervision). The objective of this joint team is to evaluate regulation policies of urban train systems, to assess their robustness to perturbations and failures, to design more efficient regulation policies and finally to provide decision support for human regulators. The project started in march 2014. A second phase of the project started in 2016, for a duration of three years. This covers in particular the CIFRE PhD of Karim Kecir.

Joint Nokia Bell Labs - Inria research lab: Several members of the team are involved in the joint research lab of Nokia Bell Labs and Inria. This lab is co-directed by Éric Fabre (Inria) and Olivier Audouin (Bell Labs), and funds joint research teams over a period of 4 years. The 3rd phase of the lab is in preparation, and 6 new joint teams will be launched in the first quarter of 2017. Sumo is involved in the proposal *Softwarization of Everything* that aims at developing techniques for the programmability, the verification and the management of software-defined networks (SDN). This covers in particular the CIFRE PhD of Arij El Majed, to start in January 2017, on the topic of Root cause analysis in reconfigurable dynamic systems.

Joint Orange Labs - Inria research lab: Éric Fabre takes part to the joint research lab of Orange Labs and Inria. This lab funds around 5 new PhD grants every year. This covers in particular the CIFRE PhD of Sihem Cherrared on the topic of Fault management in multi-tenant programmable networks.

9. Partnerships and Cooperations

9.1. National Initiatives

9.1.1. ANR

ANR STOCH-MC: Model-Checking of Stochastic Systems using approximated algorithms, 2014-2018, http://perso.crans.org/~genest/stoch.htmlweb site.
Led by SUMO.
Partners: Inria Project Team CONTRAINTES (Rocquencourt), LaBRI (Bordeaux), and LIAFA (Paris).
The aim of STOCH-MC is to perform model-checking of large stochastic systems, using controlled approximations. Two formalisms will be considered: Dynamic Bayesian Networks, which represent compactly large Markov Chains; and Markov Decision Processes, allowing non deterministic choices on top of probabilities.

ANR HeadWorks: Human-Centric Data-oriented WORKflows , 2016-2020
Led by Université Rennes 1.
Partners: Inria Project Team VALDA (LSV and ENS-ULM), Univestité Rennes 1 (DRUID), Inria SUMO, Inria Lille (LINKs), MNHN, Foule Factory.
Headwork was accepted in 2016. Participants : Loïc Hélouët, Éric Badouel.

Partners: IRISA (DRUID), ENS ULM (VALDA), Inria SUMO, Inria Lille (LINKs), MNHN, Foule Factory.

The objective of this project is to develop techniques to facilite development, deployment, and monitring of crowd-based participative applications. This requires handling complex workflows with multiple participants, incertainty in data collections, incentives, skills of contributors, ... To overcome these challenges, Headwork will define rich workflows with multiple participants, data and knowledge models to capture various kind of crowd applications with complex data acquisition tasks and human specificities. We will also address methods for deploying, verifying, optimizing, but also monitoring and adapting crowd- based workflow executions at run time.

9.1.2. IPL HAC SPECIS

The Inria Project Lab HAC SPECIS (High-performance Application and Computers, Studying PErformance and Correctness In Simulation, 2016-2020: http://hacspecis.gforge.inria.fr/) is a transversal project internal to Inria. The goal of the HAC SPECIS project is to answer the methodological needs raised by the recent evolution of HPC architectures by allowing application and runtime developers to study such systems both from the correctness and performance point of view. Inside this project, we collaborate with Martin Quinson (Myriads team) on the dynamic formal verification of high performance runtimes and applications. The PhD of The Anh Pham is granted by this project.

Partners: Inria teams AVALON (Lyon), POLARIS (Grenoble), HIEPACS, STORM (Bordeaux), MEXICO (Paris), MYRIADS, SUMO (Rennes), VERIDIS (Nancy).

Participants: Thierry Jéron, The Anh Pham.

9.1.3. National informal collaborations

The team collaborates with the following researchers:

- Yliès Falcone (CORSE LIG/Inria team in Grenoble) and Antoine Rollet (Labri Bordeaux) on the enforcement of timed properties,
- Arnaud Sangnier (IRIF) on the parameterized verification of probabilistic systems,
- Béatrice Bérard (LIP6) and Serge Haddad (LSV) on problems of opacity and diagnosis.
- Thomas Chatain, on problems related to concurrency and time,
- Eric Rutten and Gwenael delaval on the control of reconfigurable systems as well as making the ling between Reax and Heptagon / BZR (http://bzr.inria.fr/),
- Patricia Bouyer (LSV, ENS Cachan) on the analysis of probabilistic timed systems and quantitative aspects of verification,
- François Laroussinie (IRIF, UP7-Diderot) on logics for multi-agent systems.

9.2. European Initiatives

9.2.1. FP7 & H2020 Projects

Nicolas Markey is a member of Project ERC EQualIS whose principal investigator is Patricia Bouyer from LSV.

9.3. International Initiatives

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

9.3.1.1. QuantProb

Title: Quantitative analysis of non-standard properties in probabilistic models

International Partner (Institution - Laboratory - Researcher):

Technical University of Dresde (Germany) - Saxe - Christel Baier

Start year: 2016

See also: http://www.irisa.fr/sumo/QuantProb/

Quantitative information flow and fault diagnosis share two important characteristics: quantities (in the description of the system as well as in the properties of interest), and users partial knowledge. Yet, in spite of their similar nature, different formalisms have been proposed. Beyond these two motivating examples, defining a unified framework can be addressed by formal methods. Formal methods have proved to be effective to verify, diagnose, optimize and control qualitative properties of dynamic systems. However, they fall short of modelling and mastering quantitative features such as costs, energy, time, probabilities, and robustness, in a partial observation setting. This project proposal aims at developing theoretical foundations of formal methods for the quantitative analysis of partially observable systems.

9.3.2. Inria International Partners

9.3.2.1. Informal International Partners

The team collaborates on runtime enforcement with the group of Prof. Stavros Tripakis (http://users.ics.aalto.fi/stavros/) at Aalto University (Finland), where our former PhD student Srinivas Pinisetty is doing a Post-doc and with Thomas Brihaye (University of Mons) on the analysis of probabilistic timed systems.

The team has well-established collaborations with several institutes in India. CMI (Chennai Mathematical Institute, M. Mukund and N.K. Kumar), IIT Bombay (S. Akshay).

The team is building a new collaboration with Ecole Polytechnique Montreal (J. Mullins).

9.4. International Research Visitors

9.4.1. Visits of International Scientists

L. Ricker visited the SUMO team for 2 months in May-June 2016.

9.4.1.1. Internships

Robert Nsaibirni from the University of Yaoundé I joined the team from Sept. 2016 in the context of an Eiffel grant.

9.4.2. Visits to International Teams

9.4.2.1. Research Stays Abroad

Nathalie Bertrand spent a month at the Simons Institute for the theory of computing, UC Berkeley, California. She participated to the program Logical Structure in Computation (https://simons.berkeley.edu/programs/logic2016).

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events Organisation

10.1.1.1. Member of the Organizing Committees

Hervé Marchand is member of the IFAC Technical Committees (TC 1.3 on Discrete Event and Hybrid Systems) since 2005. He is member of the steering committee of MSR (Modélisation de systèmes réactifs).

Thierry Jéron and Nicolas Markey are members of the steering committee of the european summer school MOVEP (Modélisation et Vérification des Systèmes Parallèles). Nicolas Markey was co-chair of the edition that took place in Genova in July 2016.

Thierry Jéron is member of the steering committee of FMF 2017 (Formal Methods Forum) held in Toulouse in January 2017.

10.1.2. Scientific Events Selection

10.1.2.1. Chair of Conference Program Committees

Éric Badouel was Chair of conference program committee of CARI 2016.

10.1.2.2. Member of the Conference Program Committees

Éric Badouel was a member of the programme committee of ATAED 2016.

Nathalie Bertrand served on the Program Committees of the international conferences STACS'16, TACAS'16, Concur'16 and QEST'16.

Loïc Hélouët was member of the program committees of ACSD 2016 (Approaches of Concurrency for Systems Design) and SAM 2016 (System Analysis and Modeling).

Thierry Jéron served on the Program Committees of the following international conferences: ICTSS'16, RV'16, SAC-SVT 2017.

10.1.2.3. Reviewer

Nicolas Markey was reviewer for STACS 2017 and AAAI 2017.

Éric Badouel was reviewer for LICS 2016, VeCos 2016, CARI 2016, TACAS 2016, and ATAED 2016.

Loïc Hélouët was reviewer for SAM'2016, ACSD'2016, DNS'2016, STACS'2016, and ICTAC'2016

Thierry Jéron was reviewer for IEEE CASE & ISAM, CONCUR'16.

10.1.3. Journal

10.1.3.1. Member of the Editorial Boards

Éric Badouel is co-Editor-in-Chief of ARIMA Journal (https://arima.episciences.org/).

10.1.3.2. Reviewer - Reviewing Activities

Éric Fabre was reviewer for IEEE TAC, Automatica, JDEDS, CDC, and JONS.

Hervé Marchand was reviewer for JDEDS and Automatica.

Nathalie Bertrand was reviewer for JACM and JCSS.

Nicolas Markey was reviewer for FMSD and TCS.

Éric Badouel was reviewer for Fundamenta Informaticae and Mathematical review-AMS (MathSciNet).

Loïc Hélouët was reviewer for FAOC, TCS, TECS and Fundamenta Informaticae. He also served as reviewer for Mathematical review-AMS (MathSciNet).

Thierry Jéron was reviewer for FAOC and TECS.

10.1.4. Invited Talks

Nathalie Bertrand was invited speaker at MFPS international conference, and gave a lecture at MOVEP summer school.

Éric Badouel was invited speaker at VeCos 2016.

10.1.5. Scientific Expertise

Thierry Jéron served for the expertise of ANR and ASTRID (ANR/DGA) projects.

10.1.6. Research Administration

Éric Fabre is co-director, with Olivier Audouin, of the joint research lab of Nokia Bel Labs and Inria. He is member of the scientific board of the joint lab of Alstom Transport and Inria and member of the Bureau of the Scientific Board of Inria Rennes Bretagne Atlantique.

Hervé Marchand is chairman of the CUMI in Rennes.

Nathalie Bertrand is a nominated member of CNU27 (Conseil National des Universités, section 27).

Éric Badouel is co-director with Moussa Lo (UGB, Saint-Louis du Sénégal) of LIRIMA, the Inria International Lab for Africa. He is scientific officer for the African and Middle-East region at Inria European and International Partnerships Department and member of the executive board of GIS SARIMA.

Loïc Hélouët, Nathalie Bertrand and Ocan Sankur organize the weekly seminar 68NQRT at IRISA (40 talks each year).

Loïc Hélouët was elected representant of rank B researchers in the *Comité de Centre* of Inria Rennes. He is also part of the bureau of the *Comité de Centre*. He leads the P22 projects with Alstom transports and is responsible for Workpackage 2 of the Headwork ANR.

Thierry Jéron is Member Committee Substitute for COST IC1402 ARVI (Runtime Verification beyond Monitoring). He is member of the IFIP Working Group 10.2 on Embedded Systems. He is member of the COS Prospective of Irisa Rennes and member of the *Comité de Centre* of Inria Rennes. Since 2016 he is *référent chercheur* for the Inria Rennes research center.

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

Éric Fabre

Master: "ASR: introduction to distributed systems and algorithms," 12h (eq. TD), M2, Univ. Rennes 1, France.

Master: "Information theory", 30h (eq. TD), M1, Ecole Normale Superieure de Rennes, France.

Nathalie Bertrand

Licence: "Algorithmics", 18h (eq. TD), L3, Univ. Rennes 1, France.

Master: "Prépa. Agreg.", 40h (eq. TD), Ecole Normale Superieure de Rennes, France.

Loïc Hélouët

Licence: JAVA and algorithmics, L2, 40h, INSA de Renne, France.

Licence : practical studies (development of a small project), 8h, INSA de Renne, France.

Master: "Prépa. Agreg.", 8h (eq. TD)+ mock exams, Ecole Normale Superieure de Rennes, France.

10.2.2. Supervision

- PhD in progress: Engel Lefaucheux, *Controlling information in Probabilistic Systems, Sept. 2015, Nathalie Bertrand, Serge Haddad (LSV, Cachan).*
- *PhD in progress: Karim Kecir,* Régulation et robustesse des systèmes ferroviaires urbains, May 2018, Loïc Hélouët and Pierre Dersin (Alstom).
- PhD in progress: The Anh Pham, *Dynamic Formal Verification of High Performance Runtimes and Applications, Nov. 2016, Thierry Jéron, Martin Quinson (Myriads, Inria Rennes).*
- *PhD in progress: Hugo Bazille,* Diagnosability and opacity analysis of large scale systems, Oct. 2016, Blaise Genest, Éric Fabre.
- PhD in progress: Sihem Cherrared, *Fault management in multi-tenant programmable networks, Oct. 2016, Éric Fabre, Gregor Goessler (Inria Grenoble), Sofiane Imadali (Orange Labs).*

10.2.3. Juries

Éric Fabre was reviewer in the PhD defense committee of Yoann Geoffroy, *A general framework for causality analysis based on traces, for composite systems*, Dec. 2016, Univ. Grenoble Alpes. He was also jury member for the Habilitation defense of Blaise Genest, *Taming Concurrency Using Representatives, March 2016, Univ. Rennes 1.*

Hervé Marchand was member of the PhD defences of Hassan Ibrahim, Analyse à base de SAT de la diagnosticabilité et de la prédictabilité des systèmes à événements discrets centralisés et distribués (Université Paris-Sud, Gif-sur-Yvette), December 2016 and of Toussaint Tigori, *Méthodes de génération d'exécutifs temps réel (Ecole centrale de Nantes, Nantes), in November 2016.*

Nicolas Markey was reviewer in the PhD defense committee of Thanh-Tung Tran (LaBRI; supervised by Igor Walukiewicz and Frédéric Herbreteau).

10.3. Popularization

Nathalie Bertrand gave an introductory talk on graph theory and its use to solve practical problems, to grad school students following the ISN (Introduction aux Sciences du Numérique) courses.

11. Bibliography

Major publications by the team in recent years

[1] S. Akshay, B. Genest, L. Hélouët, S. Yang.*Regular Set of Representatives for Time-Constrained MSC Graphs*, in "Information Processing Letters", 2012, vol. 112, n° 14-15, p. 592-598, http://hal.inria.fr/hal-00879825.

[2] E. Badouel, M. A. Bednarczyk, A. M. Borzyszkowski, B. Caillaud, P. Darondeau.*Concurrent Secrets*, in "Discrete Event Dynamic Systems", 2007, vol. 17, n° 4, p. 425-446.

[3] E. Badouel, L. Bernardinello, P. Darondeau.*Petri Net Synthesis*, Springer, 2015, http://dx.doi.org/10.1007/978-3-662-47967-4.

[4] A. Benveniste, E. Fabre, S. Haar, C. Jard.*Diagnosis of Asynchronous Discrete Event Systems: A Net Unfolding Approach*, in "IEEE Transactions on Automatic Control", November 2003, vol. 48, n° 5, p. 714-727, RNRT project MAGDA [*DOI :* 10.1109/TAC.2003.811249], http://hal.inria.fr/inria-00638224.

[5] N. Bertrand, B. Genest, H. Gimbert.*Qualitative Determinacy and Decidability of Stochastic Games with Signals*, in "Proceedings of LICS'09", Los Angeles, États-Unis, August 2009, http://hal.archives-ouvertes.fr/hal-00356566.

[6] N. Bertrand, T. Jéron, A. Stainer, M. Krichen.*Off-line Test Selection with Test Purposes for Non-Deterministic Timed Automata*, in "Logical Methods in Computer Science", October 2012, vol. 8, n° 4:8, p. 1-33, http://hal.inria.fr/hal-00744074.

[7] J. Dubreil, P. Darondeau, H. Marchand.*Supervisory Control for Opacity*, in "IEEE Transactions on Automatic Control", May 2010, vol. 55, n° 5, p. 1089-1100 [*DOI :* 10.1109/TAC.2010.2042008].

[8] E. Fabre, A. Benveniste.*Partial Order Techniques for Distributed Discrete Event Systems: why you can't avoid using them*, in "Journal of Discrete Events Dynamical Systems", 2007, vol. 17, n° 3, p. 357-403.

[9] E. FABRE.*Trellis Processes: a Compact Representation for Runs of Concurrent Systems*, in "Journal of Discrete Event Dynamical Systems", 2007, vol. 17, n^o 3, p. 267-306.

[10] E. FABRE, L. JEZEQUEL.*Distributed optimal planning: an approach by weighted automata calculus*, in "CDC", 2009, p. 211-216.

[11] B. GAUDIN, H. MARCHAND.*An Efficient Modular Method for the Control of Concurrent Discrete Event Systems: A Language-Based Approach*, in "Discrete Event Dynamic System", 2007, vol. 17, n^o 2, p. 179-209.

[12] T. GAZAGNAIRE, B. GENEST, L. HÉLOUËT, P. THIAGARAJAN, S. YANG.*Causal Message Sequence Charts*, in "Theoretical Computer Science", 2009, 38, EA DST, http://hal.inria.fr/inria-00429538.

[13] C. JARD, T. JÉRON.*TGV: theory, principles and algorithms*, in "STTT", 2005, vol. 7, n^o 4, p. 297-315.

[14] B. JEANNET, T. JÉRON, V. RUSU, E. ZINOVIEVA.*Symbolic Test Selection Based on Approximate Analysis*, in "TACAS", Edinburgh, Royaume-Uni, 2005, http://hal.inria.fr/inria-00564617.

Publications of the year

Articles in International Peer-Reviewed Journal

[15] S. AKSHAY, L. HÉLOUËT, C. JARD, P.-A. REYNIERS.*Robustness of Time Petri Nets under Guard Enlargement*, in "Fundamenta Informaticae", 2016, vol. 143, n^o 3-4, https://hal.inria.fr/hal-01379431.

[16] E. BADOUEL, C. MORVAN, L. HÉLOUËT.*Petri Nets with Structured Data*, in "Fundamenta Informaticae", 2016, vol. 146, n^o 1, 119, https://hal.inria.fr/hal-01379245.

[17] N. BERTHIER, É. RUTTEN, N. DE PALMA, S. M.-K. GUEYE.*Designing Autonomic Management Systems by using Reactive Control Techniques*, in "IEEE Transactions on Software Engineering", July 2016, vol. 42, n^o 7, 18, https://hal.inria.fr/hal-01242853.

[18] R. BRENGUIER, J.-F. RASKIN, O. SANKUR.*Assume-admissible synthesis*, in "Acta Informatica", 2016 [*DOI :* 10.1007/S00236-016-0273-2], https://hal.inria.fr/hal-01373538.

[19] B. BÉRARD, L. HÉLOUËT, J. MULLINS.*Non-interference in partial order models*, in "ACM Transactions on Embedded Computing Systems (TECS)", 2016, https://hal.inria.fr/hal-01379451.

[20] Y. FALCONE, T. JÉRON, H. MARCHAND, S. PINISETTY.*Runtime Enforcement of Regular Timed Properties by Suppressing and Delaying Events*, in "Science of Computer Programming", March 2016 [*DOI :* 10.1016/J.SCICO.2016.02.008], https://hal.inria.fr/hal-01281727.

[21] S. JACOBS, R. BLOEM, R. BRENGUIER, R. EHLERS, T. HELL, R. KÖNIGHOFER, G. A. PÉREZ, J.-F. RASKIN, L. RYZHYK, O. SANKUR, M. SEIDL, L. TENTRUP, A. WALKER.*The first reactive synthesis competition (SYNTCOMP 2014)*, in "International Journal on Software Tools for Technology Transfer", April 2016 [*DOI :* 10.1007/S10009-016-0416-3], https://hal.inria.fr/hal-01373547.

Invited Conferences

[22] R. BRENGUIER, L. CLEMENTE, P. HUNTER, G. A. PÉREZ, M. RANDOUR, J.-F. RASKIN, O. SANKUR, M. SASSOLAS.*Non-Zero Sum Games for Reactive Synthesis*, in "LATA 2016 : 10th International Conference on Language and Automata Theory and Applications", Prague, Czech Republic, A.-H. DEDIU, J. JANOUŠEK, C. MARTÍN-VIDE, B. TRUTHE (editors), Lecture Notes in Computer Science, Springer, March 2016, vol. 9618, p. 3-23 [*DOI :* 10.1007/978-3-319-30000-9_1], https://hal.inria.fr/hal-01373546.

International Conferences with Proceedings

[23] S. AKSHAY, B. GENEST, L. HÉLOUËT.*Decidable Classes of Unbounded Petri Nets with Time and Urgency*, in "Application and Theory of Petri Nets and Concurrency - 37th International Conference, PETRI NETS, 2016", Torun, Poland, F. KORDON, D. MOLDT (editors), Application and Theory of Petri Nets and Concurrency - 37th International Conference, PETRI NETS, 2016, Springer, June 2016, n^{o} 9698, p. 301 - 322 [*DOI :* 10.1007/978-3-319-39086-4_18], https://hal.inria.fr/hal-01379414.

[24] S. AKSHAY, B. GENEST, B. KARELOVIC, N. VYAS.*On Regularity of unary Probabilistic Automata*, in "STACS 2016", Orléans, France, STACS 2016, 2016, https://hal.archives-ouvertes.fr/hal-01245037.

[25] S. AKSHAY, L. HÉLOUËT, R. PHAWADE.*Combining Free Choice and Time in Petri Nets*, in "23rd International Symposium on Temporal Representation and Reasoning", Lyngby, Denmark, IEEE (editor), 23rd International Symposium on Temporal Representation and Reasoning, Technical University of Denmark , October 2016, https://hal.inria.fr/hal-01379440.

[26] N. BERTRAND, P. BOUYER, T. BRIHAYE, P. CARLIER.*Analysing Decisive Stochastic Processes*, in "43rd International Colloquium on Automata, Languages, and Programming (ICALP 2016)", Rome, Italy, LiPICS, 2016, vol. 55, 14 [*DOI :* 10.4230/LIPIcs.ICALP.2016.101], https://hal.inria.fr/hal-01397794.

[27] N. BERTRAND, S. HADDAD, E. LEFAUCHEUX.*Accurate approximate diagnosability of stochastic systems*, in "10th International Conference on Language and Automata Theory and Applications", Prague, Czech Republic, Springer, March 2016, https://hal.inria.fr/hal-01220954.

[28] N. BERTRAND, S. HADDAD, E. LEFAUCHEUX.*Diagnosis in Infinite-State Probabilistic Systems* , in "27th International Conference on Concurrency Theory (Concur 2016)", Québec city, Canada, 27th International Conference on Concurrency Theory (Concur 2016), August 2016 [*DOI :* 10.4230/LIPIcs.CONCUR.2016.37], https://hal.inria.fr/hal-01373354.

[29] R. BRENGUIER, G. A. PÉREZ, J.-F. RASKIN, O. SANKUR.*Admissibility in Quantitative Graph Games*, in "36th IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science", Chennai, India, December 2016, https://hal.inria.fr/hal-01373542.

[30] L. HÉLOUËT, K. KECIR.*Realizability of Schedules by Stochastic Time Petri Nets with Blocking Semantics*, in "37th International Conference, PETRI NETS 2016", Torun, Poland, F. KORDON, D. MOLDT (editors), Application and Theory of Petri Nets and Concurrency, Fabrice Kordon and Daniel Moldt, June 2016, n^{o} 9698, p. 155-175 [*DOI :* 10.1007/978-3-319-39086-4_11], https://hal.inria.fr/hal-01379424.

[31] R. F. J. NSAIBIRNI, E. BADOUEL, G. TEXIER, G.-E. KOUAMOU.*Active-Workspaces: A Dynamic Collaborative Business Process Model for Disease Surveillance Systems*, in "Worldcomp'16- The 2nd International Conference on Health Informatics and Medical Systems", Las Vegas, United States, July 2016, https://hal.inria.fr/hal-01323561.

[32] S. K. PALANIAPPAN, M. PICHENÉ, G. BATT, E. FABRE, B. GENEST.*A Look-Ahead Simulation Algorithm for DBN Models of Biochemical Pathways*, in "Hybrid Systems Biology, 5th International Workshop, HSB 2016", Grenoble, France, E. CINQUEMANI, A. DONZÉ (editors), Lecture Notes in Computer Science, October 2016, vol. 9957, p. 3-15 [*DOI :* 10.1007/978-3-319-47151-8_1], https://hal.inria.fr/hal-01406115.

[33] S. PINISETTY, V. PREOTEASA, S. TRIPAKIS, T. JÉRON, Y. FALCONE, H. MARCHAND.*Predictive Runtime Enforcement **, in "SAC 2016 31st ACM Symposium on Applied Computing", Pisa, Italy, ACM, April 2016, 6 [*DOI :* 10.1145/2851613.2851827], https://hal.inria.fr/hal-01244369.

[34] L. S. L. RICKER, H. MARCHAND.*Finding the weakest link(s): Coalition games for decentralized discrete-event control*, in "55th IEEE Conference on Decision and Control", Las-Vegas, United States, December 2016, https://hal.inria.fr/hal-01373709.

Conferences without Proceedings

[35] E. FABRE, L. HÉLOUËT, E. LEFAUCHEUX, H. MARCHAND.*Diagnosability of Repairable Faults*, in "13th International Workshop on Discrete Event Systems", Xi'an, China, 2016, p. 256-262, (Version Longue), https://hal.inria.fr/hal-01302562.

[36] O. SANKUR, J.-P. TALPIN.*An Abstraction Technique For Parameterized Model Checking of Leader Election Protocols: Application to FTSP*, in "23rd International Conference on Tools and Algorithms for the Construction and Analysis of Systems (TACAS)", Uppsala, Sweden, April 2017, https://hal.archives-ouvertes.fr/hal-01431472.

Books or Proceedings Editing

[37] M. LO, E. BADOUEL, N. GMATI (editors). *Proceedings of CARI 2016*, July 2016, 513, https://hal.inria.fr/hal-01350039.

Research Reports

[38] N. BERTRAND, S. HADDAD, E. LEFAUCHEUX.*Diagnosis in Infinite-State Probabilistic Systems (long version)*, Inria Rennes ; LSV, ENS Cachan, June 2016, https://hal.inria.fr/hal-01334218.

[39] P. BOUYER, P. HOFMAN, N. MARKEY, M. RANDOUR, M. ZIMMERMANN.*Bounding Average-energy Games*, ArXiv, October 2016, n^{o} 1610.07858, https://hal.archives-ouvertes.fr/hal-01410144.

[40] P. BOUYER, V. JUGÉ, N. MARKEY.*Dynamic Complexity of Parity Games with Bounded Tree-Width*, ArXiv, October 2016, https://hal.archives-ouvertes.fr/hal-01410169.

Other Publications

[41] B. BÉRARD, L. HÉLOUËT, J. MULLINS.*Non-interference in partial order models*, February 2016, working paper or preprint, https://hal.inria.fr/hal-01280043.

[42] L. HÉLOUËT, K. KECIR.*Realizability of Schedules by Stochastic Time Petri Nets with Blocking Semantics (regular paper)*, March 2016, working paper or preprint, https://hal.inria.fr/hal-01284682.

References in notes

[43] P. A. ABDULLA, A. NYLÉN.*Timed Petri Nets and BQOs*, in "ICATPN", LNCS, 2001, vol. 2075, p. 53-70.

[44] E. BADOUEL, L. HÉLOUËT, C. MORVAN, R. F. J. NSAIBIRNI.*Active Workspace: Distributed Collaborative Systems based on Guarded Attribute Grammars*, in "ACM SIGAPP Applied Computing Review", 2015, vol. 6, n^{o} 33, p. 6–33.

[45] B. BÉRARD, L. HÉLOUËT, J. MULLINS.*Non-interference in Partial Order Models*, in "Proc. of 15th Int. Conf. on Application of Concurrency to System Design, ACSD 2015", IEEE Computer Society, 2015, p. 80–89.

[46] N. D. JONES, L. H. LANDWEBER, Y. E. LIEN.*Complexity of Some Problems in Petri Nets*, in "Theor. Comput. Sci.", 1977, vol. 4, n^{o} 3, p. 277–299.

[47] P. M. MERLIN.*A Study of the Recoverability of Computing Systems*, University of California, Irvine, CA, USA, 1974.

[48] A. PURI.*Dynamical Properties of Timed Automata*, in "In DEDS", 2000, vol. 10, n^{o} 1-2, p. 87–113.

[49] V. V. RUIZ, F. C. GOMEZ, D. DE FRUTOS-ESCRIG.*On Non-Decidability of Reachability for Timed-Arc Petri Nets*, in "PNPM", IEEE Computer Society, 1999, p. 188–.

[50] B. WALTER.*Timed Petri-Nets for Modelling and Analysing Protocols with Real-Time Characteristics*, in "Proc. of PSTV", 1983, p. 149-159.

Team TACOMA

TAngible COMputing Architectures

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

RESEARCH CENTER
Rennes - Bretagne-Atlantique

THEME
Distributed programming and Software engineering

Table of contents

Team TACOMA

Creation of the Team: 2014 January 01

Keywords:

Computer Science and Digital Science:

1.2. - Networks
1.2.5. - Internet of things
1.2.6. - Sensor networks
1.2.7. - Cyber-physical systems
1.3. - Distributed Systems
1.4. - Ubiquitous Systems
2.3.2. - Cyber-physical systems
2.5. - Software engineering
2.6. - Infrastructure software
2.6.1. - Operating systems
2.6.2. - Middleware
4.8. - Privacy-enhancing technologies
5.11. - Smart spaces
5.11.1. - Human activity analysis and recognition
5.11.2. - Home/building control and interaction

Other Research Topics and Application Domains:

3.1. - Sustainable development
3.1.1. - Resource management
4.4. - Energy delivery
4.4.1. - Smart grids
4.5.2. - Embedded sensors consumption
6.1. - Software industry
6.1.1. - Software engineering
6.1.2. - Software evolution, maintenance
6.2.2. - Radio technology
6.3.3. - Network Management
6.4. - Internet of things
6.6. - Embedded systems
7.2. - Smart travel
7.2.1. - Smart vehicles
7.2.2. - Smart road
8.1. - Smart building/home
8.2. - Connected city
8.5.2. - Crowd sourcing
8.5.3. - Collaborative economy
9.8. - Privacy
9.10. - Ethics

TACOMA ends in December 2016. Since September 2016, the team's composition has evolved including Jean-Marie Bonnin from Telecom Bretagne, and TACOMA addresses new applications domains (mobility, smart cities). The presentation of the research program in this report has been revised to reflect these recent evolutions. This expansion of the team's skill set will be helpful in the future to tackle, at different levels (sensors, network, software), architectural concerns that impairs the deployment of smarter environments.

1. Members

Research Scientist
Paul Couderc [Inria, Researcher]

Faculty Members
Frédéric Weis [Team leader, Univ. Rennes I, Associate Professor, HDR]
Jean-Marie Bonnin [Telecom Bretagne, Professor, from Sept 2016, HDR]
Yoann Maurel [Univ. Rennes I, Associate Professor]

Technical Staff
Michele Dominici [Univ. Rennes I]
Nebil Ben Mabrouk [Univ. Rennes I, until Jul 2016]

PhD Students
Adrien Capaine [OyaLight, granted by CIFRE]
Indra Ngurah [Telecom Bretagne, from May 2016]
Alexandre Rio [OKWIND, granted by CIFRE, from Oct 2016]
Liouane Zaineb [Co-tutelle Univ. Rennes 1 with Univ. Monastir]

Visiting Scientist
Christophe Couturier [Telecom Bretagne, from Sept 2016]

Administrative Assistant
Fabienne Cuyollaa [Inria]

2. Overall Objectives

2.1. Overall Objectives

The technologies necessary for the development of pervasive applications are now widely available and accessible for many uses: short/long-range and low energy communications, a broad variety of visible (smart objects) or invisible (sensors and actuators) objects, as well as the democratization of the Internet of Things (IoT). Large areas of our living spaces are now instrumented. The concept of Smart Spaces is about to emerge, based upon both massive and apposite interactions between individuals and their everyday working and living environments: residential housing, public buildings, transportation, etc. The possibilities of new applications are boundless. Many scenarios have been studied in laboratories for many years and, today, a real ability to adapt the environment to the behaviors and needs of users can be demonstrated. However mainstream pervasive applications are barely existent, at the notable exception of the ubiquitous GPS-based navigators. The opportunity of using vast amount of data collected from the physical environments for **several application domains** is still largely untapped. The applications that interact with users and act according to their environment with a large autonomy are still very specialized. They can only be used in the environment they had especially been developed for (for example "classical" home automation tasks: comfort, entertainment, surveillance). They are difficult to adapt to increasingly complex situations, even though the environments in which they evolve are more open, or change over time (new sensors added, failures, mobility etc.).

Developing applications and services that are ready to deploy and evolve in different environments should involve significant cost reduction. Unfortunately, designing, testing and ensuring the maintenance as well as the evolution of a pervasive application remains very complex. In our view, the lack of resources by which properties of the real environment are made available to application developers is a major concern. Building a pervasive application involves implementing one or more logical control loops which include four stages (see figure 1-a): (1) data collection in the real environment, (2) the (re)construction of information that is meaningful for the application and (3) for decision making, and finally, (4) action within the environment. While many decision-algorithms have been proposed, the **collection** and **construction** of a reliable and relevant perception of the environment and, in return, **action** mechanisms within the environment still pose major challenges that the TACOMA project is prepared to deal with.

Most current solutions are based on a massive collection of raw data from the environment, stored on remote servers. Figure 1-a illustrates this type of approach. Exposure of raw sensor values to the decision-making process does not allow to build relevant contexts that a pervasive application actually needs in order to shrewdly act/react to changes in the environment. So, the following is left up to the developer:

- To characterize more finely raw data beyond its simple value, for example, the acquisition date, the nature of network links crossed to access the sensor, the durability and accuracy of value reading, etc.
- To exploit this raw data to calculate a relevant abstraction for the application, such as, whether the room is occupied, or whether two objects are in the same physical vicinity.
- To modify the environment when possible.

Traditional software architectures isolate the developer from the real environment that he oftentimes has to depict according to complex, heavy and expensive processes. However, objects and infrastructure integrated into user environments could provide a more suitable support to pervasive applications: description of the actual system's state can be richer, more accurate, and, meanwhile, easier to handle; the applications' structure can be distributed by being built directly into the environment, facilitating scalability and resilience by the processing autonomy; finally, moving processing closer to the edge of the network avoids major problems of data sovereignty and privacy encountered in infrastructures very dependent on the cloud. We strongly believe in the advantages of specific approaches to the fields of **edge computing** and **fog computing**, which will reveal themselves with the development of Smart Spaces and an expansive growth of the number of connected objects. Indeed, ensuring the availability and reliability of systems that remain frugal in terms of resources will become in the end a major challenge to be faced in order to allow proximity between processing and end-users. Figure 1-b displays the principle of "using data at the best place for processing". Fine decisions can be made closer to the objects producing and acting on the data, local data characterization and local processing de-emphasize the computing and storage resources of the cloud (which can be used for example to store selected/transformed data for global historical analysis or optimization).

TACOMA aims at developing a comprehensive set of new **interaction models** and **system architectures** to considerably help pervasive application designers in the development phase with the side effect to ease the life cycle management. We follow two main principles:

- Leveraging local properties and direct interactions between objects, we would be able to enrich and to manage locally data produced in the environment. The application would then be able to build their knowledge about their environment (perception) in order to adjust their behavior (eg. level of automation) to the actual situation.
- Pervasive applications should be able to describe requirements they have on the quality of their environment perception. We would be able to achieve the minimum quality level adapting the diversity of the sources (data fusion/aggregation), the network mechanisms used to collect the data (network/link level) and the production of the raw data (sensors).

3. Research Program

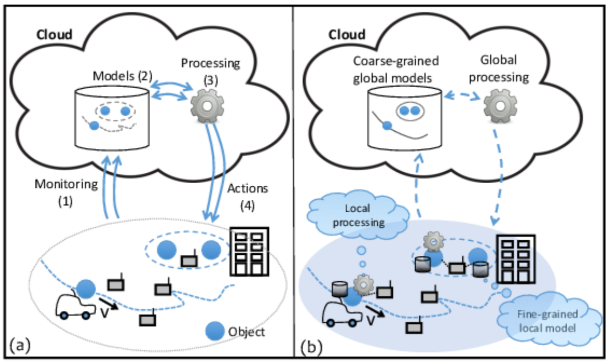


Figure 1. Adaptation processes in pervasive environments

3.1. Collecting pertinent information

In our model, applications adapt their behavior (for instance, the level of automation) to the quality of their perception of the environment. This is important to alleviate the development constraint we usually have on automated system. We "just" have to be sure a given process will always operate at the right automation level given the precision, the completeness or the confidence it has on its own perception. For instance, a car passing through a cross would choose its speed depending on the confidence it has gained during perception data gathering. When it has not enough information or when it could not trust it, it should reduce the automation level, therefore the speed, to only rely on its own sensors. Such adaptation capability shift requirements from the design and deployment (availability, robustness, accuracy, etc.) to the **assessment of the environment perception** we aim to facilitate in this first research axis.

Data characterization. The quality (freshness, accuracy, confidence, reliability, confidentiality, etc.) of the data are of crucial importance to assess the quality of the perception and therefore to ensure proper behavior. The way data is produced, consolidated, and aggregated while flowing to the consumer has an impact on its quality. Moreover part of these quality attributes requires to gather information at several communication layers from various entities. For this purpose, we want to design **lightweight cross-layer interactions** to collect relevant data. As a "frugality" principle should guide our approach, it is not appropriate to build all attributes we can imagine. It is therefore necessary to identify attributes relevant to the application and to have mechanisms to activate/deactivate at run-time the process to collect them.

Data fusion. Raw data should be directly used only to determine low-level abstraction. Further help in abstracting from low-level details can be provided by **data fusion** mechanisms. A good (re)construction of a meaningful information for the application reduces the complexity of the pervasive applications and helps the developers to concentrate on the application logic rather on the management of raw data. Moreover, the reactivity required in pervasive systems and the aggregation of large amounts of data (and its processing) are antagonists. We study **software services that can be deployed closer to the edge of the network**. The exploration of data fusion technics will be guided by different criteria: relevance of abstractions produced for pervasive applications, anonymization of exploited raw data, processing time, etc.

Assessing the correctness of the behavior. To ease the design of new applications and to align the development of new products with the ever faster standard developments, continuous integration could be used in parallel with continuous conformance and interoperability testing. We already participate in the design of new shared platforms that aims at facilitating this providing remote testing tools. Unfortunately, it is not possible to be sure that all potential peers in the surrounding have a conform behavior. Moreover, upon failure or security breach, a piece of equipment could stop to operate properly and lead to global mis-behavior. We want to propose conceptual tools for **testing at runtime devices in the environment**. The result of such conformance or interoperability tests could be stored safely in the environment by authoritative testing entity. Then application could interact with the device with a higher confidence. The confidence level of a device could be part of the quality attribute of the information it contributed to generate. The same set of tools could be used to identify misbehaving device for maintenance purpose or to trigger further testing.

3.2. Building relevant abstraction for new interactions

The pervasive applications are often designed in an ad hoc manner depending on the targeted application area. Ressources (sensors / actuators, connected objets etc.) are often used in silos which complexify the implementation of rich pervasive computing scenarios. In the second research axis, we want to get away from technical aspects identifying **common and reusable system mechanisms** that could be used in various applications.

Tagging the environment. Information relative to environment could be stored by the application itself, but it could be complex to manage for mobile application since it could cross a large number of places with various features. Moreover the developer has to build its own representation of information especially when he wants to share information with other instances of the same application or with other applications. A promising approach is to store and to maintain this information associated to an object or to a place, in the environment itself. The infrastructure should provide services to application developers: add/retrieve information in the environment, share information and control who can access it, add computed properties to object for further usage. We want to study an **extensible model to describe and augment the environment**. Beyond a simple distributed storage, we have in mind a new kind of interaction between pervasive applications and changing environment and between applications themselves.

Taking advantages of the spatial and temporal relationships. To understand the world they have to interact with, pervasive applications often have to (re)built a model of it from the exchange they have with others or from their own observations. A part of the programmer's task consists in building a model of the spatial layout of the objects in the surrounding. The term *layout* can be understood in several ways: the co-location of multiple objects in the same vicinity, the physical arrangement of two objects relative to each other, or even the crossing of an object of a physical area to another, etc. Determining remotely these spatial properties (see figure 1-a) is difficult without exchanging a lot of information. Properties related to the spatial layout are far easier to characterize locally. They could be abstracted from interaction pattern without any complex virtual representation of the environment (see figure 1-b). We want to be able to rely on this type of spatial layout in a pervasive environment. In the prior years, the members of TACOMA already worked on **models for processing object interactions** in the physical world to automatically trigger processing. This was the case in particular of the spatial programming principle: physical space is treated as a tuple-space in which objects are automatically synchronized according to their spatial arrangement. We want to follow this approach by considering **richer and more expressive programming models.**

3.3. Acting on the environment

The conceptual tools we aim to study must be *frugal*: they use as less as possible resources, while having the possibility to use much more when it is required. Data needed by an application are not made available for "free"; for example, it costs energy to measure a characteristic of the environment, or to transmit it. So this "design frugality" requires **a fine-grained control** on how data is actually collected from the environment. The third research axis aims at designing solutions that give this control to application developers by **acting on the environment**.

Acting on the data collection. We want to be able to identify which information are reality needed during the perception elaboration process. If a piece of data is missing to build a given information with the appropriate quality level, the data collection mechanism should find relevant information in the environment or modify the way it aggregate it. These could lead to a modification of the behavior of the network layer and the path the piece of data use in the aggregation process.

Acting on object interactions. Object in the environment could adapt their behavior in a way that strongly depend on the object itself and that is difficult to generalize. Beyond the specific behaviors of actuators triggered through specialized or standard interfaces, the production of information required by an application could necessitate an adaptation at the object level (eg. calibration, sampling). The environment should then be able to initiate such adaption transparently to the application, which may not know all objects it passes by.

Adapting object behaviors. The radio communication layers become more flexible and able to adapt the way they use energy to what is really required for a given transmission. We already study how beamforming technics could be used to adapt multicast strategy for video services. We want to show how playing with these new parameters of transmissions (eg. beamforming, power, ...) allows to control spatial relationships objects could have. There is a tradeoff to find between the capacity of the medium, the electromagnetic pollution and the reactivity of the environment. We plan to expend our previous on interface selection and more generally on what we call **opportunistic networking**.

4. Application Domains

4.1. Pervasive applications in Smart Building

A Smart Building is a living space equipped with information-and-communication-technology (ICT) devices conceived to collaborate in order to anticipate and respond to the needs of the occupants, working to promote their comfort, convenience, security and entertainment while preserving their natural interaction with the environment.

The idea of using the Pervasive Computing paradigm in the Smart Building domain is not new. However, the state-of-the-art solutions only partially adhere to its principles. Often the adopted approach consists in a heavy deployment of sensor nodes, which continuously send a lot of data to a central elaboration unit, in charge of the difficult task of extrapolating meaningful information using complex techniques. This is a *logical approach*. TACOMA proposed instead the adoption of a *physical approach*, in which the information is spread in the environment, carried by the entities themselves, and the elaboration is directly executed by these entities "inside" the physical space. This allows performing meaningful exchanges of data that will thereafter need a less complicated processing compared to the current solutions. The result is a smart environment that can, in an easier and better way, integrate the context in its functioning and thus seamlessly deliver more useful and effective user services. Our contribution aims at implementing the physical approach in a smarter environment, showing a solution for improving both comfort and energy savings.

4.2. Metamorphic House

The motivation for metamorphic houses is that many countries, including France, are going through socio-demographic evolutions, like growth of life expectancy and consequent increase in the number of elderly people, urbanization and resource scarcity. Households experience financial restrictions, while housing costs increase with the raise of real estate and energy prices [5].

Important questions arise concerning the future of housing policies and ways of living. We observe novel initiatives like participative housing and developing behaviors, including house-sharing, teleworking and longer stay of children in parents' homes.

To tackle the challenges raised by these emerging phenomena, future homes will have to be modular, upgradeable, comfortable, sparing of resources. They should be integrated in the urban context and exchange information with other homes, contribute to reducing the distances to be covered daily and respect the characteristics of the territory where they are located.

To reach these goals, metamorphic domestic environments will modify their shape and behavior to support activities and changes in life cycle of occupants, increase comfort and optimize the use of resources. Thanks to Information and Communication Technologies (ICT) and adaptive building elements, the same physical spaces will be transformed for different uses, giving inhabitants the illusion of living in bigger, more adapted and more comfortable places.

4.3. Automation in Smart City

The domain of Smart Cities is still young but it is already a huge market which attract number of companies and researchers. It is also multi-fold as the words "smart city" gather multiple meanings. Among them one of the main responsibilities of a city, is to organize the transportation of goods and people. In intelligent transportation systems (ITS), ICT technologies have been involved to improve planification and more generally efficiency of journeys within the city. We are interested in the next step where efficiency would be improved locally relying on local interactions between vehicles, infrastructure and people (smartphones).

For the future autonomous vehicle are now in the spotlight, since a lot of works has been done in recent years in automotive industry as well as in academic research centers. Such unmanned vehicle could strongly impact the organisation of the transportation in our cities. However, due to the lack of a definition of what is an "autonomous" vehicle it remains still difficult to see how these vehicles will interact with their environment (eg. road, smart city, houses, grid, etc.). From augmented perception to fully cooperative automated vehicle, the autonomy covers various realities in terms of interaction the vehicle relies on. The extended perception relies on communication between the vehicle and surrounding roadside equipments. This help the driving system to build and maintain an accurate view of the environment. But at this first stage the vehicle only uses its own perception to make its decisions. At a second stage, it will take advantages of local interaction with other vehicles through car-to-car communications to elaborate a better view of its environment. Such "cooperative autonomy" does not try to reproduce the human behavior anymore, it strongly relies on communication between vehicles and/or with the infrastructure to make decision and to acquire information on the environment. Part of the decision could be centralized (almost everything for an automatic metro) or coordinated by a roadside component. The decision making could even be fully distributed but this put high constraints on the communications. Automated vehicles are just an exemple of smart city automated processes that will have to share information within the surrounding to make their decisions.

4.4. Pervasive applications in uncontrolled environnements

Some limitations of existing RFID technology become challenging: unlike standard RFID application scenarios, pervasive computing often involves uncontrolled environment for RFID, where tags and reader have to operate in much more difficult situations that those usually encountered or expected for classical RFID systems.

RFID technology is to avoid missing tags when reading multiple objects, as reading reliability is affected by various effects such shadowing or wave power absorption by some materials. The usual applications of RFID operate in a controlled environment in order to reduce the risk of missing tags while scanning objects.

In pervasive computing applications, a controlled reading environment is extremely difficult to achieve, as one of the principle is to enhance existing processes "in situ", unlike the controlled conditions that can be found in industrial processes. Consider for example a logistic application, where RFID tags could be used on items inside a package in order to check for its integrity along the shipping process. Tags would likely be placed randomly on items inside the package, and reading conditions would be variable depending on where the package is checked.

RFID operation in uncontrolled environments is challenging because RFID performance is affected by multiple parameters, in particular:

- Objects materials (on which tags are attached to),
- Materials in the surrounding environment,
- RFID frequency spectrum,
- Antenna nature and placement with respect to the tags.

In controlled environment, the difficulty to read tags can be limited by using the appropriate parameters to maximize the RFID performance for the application. But in many cases, it is needed to read large number of objects of various nature, arranged randomly in a given area or container. **Most pervasive computing applications fall in this context**.

5. New Software and Platforms

5.1. THEGAME: data fusion for Smart Home and Smart Building

KEYWORDS: Smart home - Smart building

- Participants: Yoann Maurel and Frédéric Weis
- Partner: Université de Rennes 1
- Contact: Frédéric Weis
- URL: https://github.com/bpietropaoli/THEGAME/ (C-version)
- URL: https://bitbucket.org/TACOMA-irisa/java-bft (Java-version)

SCIENTIFIC DESCRIPTION

Context-aware applications have to sense the environment in order to adapt themselves and provide with contextual services. This is the case of Smart Homes equipped with sensors and augmented appliances. However, sensors can be numerous, heterogeneous and unreliable. Thus the data fusion is complex and requires a solid theory to handle those problems. The aim of the data fusion, in our case, is to compute small pieces of context we call context attributes. Those context attributes are diverse and could be for example the presence in a room, the number of people in a room or even that someone may be sleeping in a room. For this purpose, we developed an implementation of the belief functions theory (BFT). THE GAME (THeory of Evidence in a lanGuage Adapted for Many Embedded systems) is made of a set of Libraries. It provides the basics of belief functions theory, computations are optimized for an embedded environment (binary representation of sets, conditional compilation and diverse algorithmic optimizations).

THE GAME is published under apache licence. It is maintained and experimented within a sensor network platform developed by TACOMA since June 2013.

5.2. Platform Pervasive_RFID

KEYWORDS: Composite objects - RFID

- Participants: Paul Couderc and Anthony Blair
- Partner: Université de Rennes 1 (IETR)
- Contact: Paul Couderc

SCIENTIFIC DESCRIPTION

In 2016 we completed the RFID experiment testbed realized in 2014-2015 in collaboration with IETR (see Figure 2). This system allows both interactive testing as well as long running experiments of RFID reading protocols. It comprises a software platform allowing fine control over all dynamic aspects influencing RFID readings: movements for target and antenna, RFID reader configuration, and smart antenna configuration .(diversity and power control).

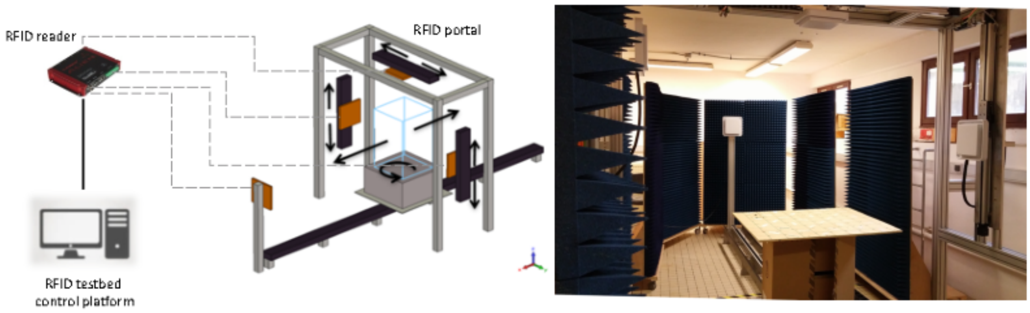


Figure 2. RFID testbed

5.3. Metamorphic Housing platform and Software

KEYWORDS: Smart Home - Metamorphic House

- Partner: Université de Rennes 1
- Partner: Université de Rennes 1 (Fondation Rennes 1)
- Contact: Michele Dominici and Frédéric Weis

SCIENTIFIC DESCRIPTION- SOFTWARE

As part of the experimentation of the On-demand room see 6.6, we have developed a software system that will be used to manage the room and provide functionalities to end users and building managers (access control, electrical and time consumption monitoring and report, room state display...). The software is expected to be deployed in the building that hosts the experimentation. This software is co-developed by Michele Dominici (University of Rennes 1), Guillermo Andrade (SED Inria) and Ghislain Nouvel (MobBI platform). Contributions might be provided by members of the Diverse project-team. Intellectual protection is expected to be applied on such software.

SCIENTIFIC DESCRIPTION- PLATFORM

In 2015, we realized a prototype of the on-demand room as an immersive interactive virtual-reality application, leveraging the Immersia platform (see https://raweb.inria.fr/rapportsactivite/RA2015/tacoma/uid29.html), with real domestic appliances connected to Immersa. In 2016, the experimentation of the On-demand room is organized in the following steps: modification of the original building to create a common, On-demand room between two apartments; deployment of the computer and hardware and software that we are developing; rental of the apartments to two households, for an estimated duration of one year. The building that will host the experimentation is showed in Figure 3. During the rental of the apartments, data will be collected and stored about the use of the room by households. Data will include time of occupation, mode (private or shared), consumptions, errors etc. The On-demand room will thus constitute an experimentation platform, where real people live and produce data that can be analyzed for statistical purposes. Produced data could also be used in combination with interviews of the occupants to improve the functionalities of the On-demand room, evaluate acceptance and appropriation.

5.4. ISO/IEC 15118-2 Open source Implementation

KEYWORDS: Smart Grid - Intelligent Transport System

- Partner: Telecom Bretagne
- Contact: Jean-Marie Bonnin

Figure 3. On-demand room real experimentation

SCIENTIFIC DESCRIPTION

The ISO/IEC 15118 standard, named "Road vehicles ? Vehicle-to-Grid Communication Interface", defines how an electric vehicle and a charging station should communicate. It enables the Smart Charging of electric vehicles by allowing them to plan their charging sessions. As we want to be able to manage the charge of electric vehicles in our micro Smart Grid systems, we decided to implement the protocol defined by this standard. The goal is also to participate actively in the design of the new version of this protocol. During a charging session the charging station provides the vehicle with the status of the electric power grid. The vehicle is then able to plan its sharing session accordingly. It sends back its charge plan to the charging station, so that the Smart Grid is aware of it. The protocol also provides security and authentication features.

This software platform was implemented onto small PCs, and was used to control the charge in a small and portable demonstration platform, to demonstrate how it is possible to interconnect this high level decision and communication software with low level components, such as a Battery Management System (BMS), and a battery charger. In 2016, in the context of the Greenfeed project our software has been demonstrated to control the charge of the electric vehicle during the final demonstration of the project. The integration work has been done in collaboration with VeDeCom.

6. New Results

6.1. RFID for pervasive computing environments

Participants: Nebil Ben Mabrouk, Frédéric Weis, Paul Couderc [contact].

Here 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 "self-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. We previously applied the self-describing objects approach to the waste management domain, which has shown to be a specially challenging situation for RFID. This challenge is found more generally in pervasive computing scenarios involving RFID reading in uncontrolled environments (see section 4.4).

We achieved the following results:

- We showed the importance of diversity in the context of challenging RFID reading. A reconfigurable antenna was designed to support dynamic reading protocols.
- A software approach based on error correcting code was developed to support robust data storage in groups of RFID.
- An innovative RFID testbed for experimenting a large range of RFID situations/applications was operational (minus some features to be completed), supported by a simulation environment and a control environment.
- A patent was filed and some contacts made with RFID companies.

However, the supports for implementing dynamic reading protocols were lacking, both on the software and the radio side. The following further progress were made:

- The four elements diversity antenna designed in first phase was implemented.
- The control software has been greatly improved. A new environment was designed, offering powerful and flexible programming capabilities for easy prototyping of RFID reading scenarios and collecting experiments results. A simulator of the testbed was also developed, allowing off-site developments. This work is supported by the RFID-Lab ADT.
- Motion-induced improvements of RFID reliability were experimented, as shown below in Figure 4.
- A significant dissemination efforts toward the industry was made, and we have good hope that some of the contacts will lead to perspectives.

An example of motion-assisted RFID readings implemented is shown in figure 4: a matrix of 32 RFID tags are arranged in reduced power conditions, so that the tags are near their sensivity limit. In such conditions, 20% of the tags failed to be read by the reader. By coupling the reading with a rotation of 210 deg, we show that all the missing tags are progressively recovered.

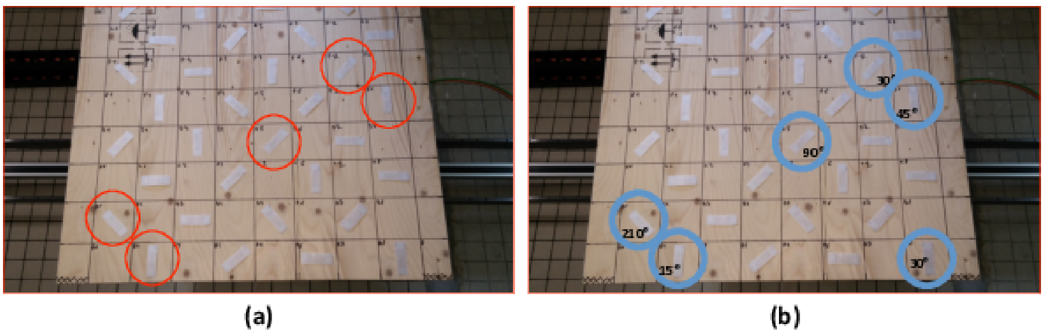


Figure 4. (a) initial read, 20% of the tags are missing (b) After 210 deg of rotation, all the tags are recovered

6.2. Building an extensible information sharing mechanism

Participants: Adrien Capaine, Yoann Maurel, Frédéric Weis [contact].

Context aware applications adapt their behavior based on information they can collect on their surrounding environment. Most of these data are provided by third-party software, sensors or computed by the application itself. A striking challenge facing the building of comprehensive pervasive system is the lack of integration between the different services provided by third parties. In this project, we intend to study and to provide mechanisms to enhance information sharing between applications and more specifically to augment information on the surrounding environment. The idea is to endow applications with the capability to increase or augment information on the physical world they are interacting with and to retrieve and share these data seamlessly depending on their location. Such mechanism aims at providing a complementary source of information in order to improve the process of choosing the best service/information provider and to help them keeping additional information on physical resources such as environment specific configuration (e.g., calibration data).

The idea of augmenting information on the physical world is not new. It has been done for centuries in the real world. For instance, the Little Thumb sowed pebbles to find his way just as hikers use cairns so as not to get lost. In daily life, people use sticky notes on pieces of hardware or objects to keep relevant information on their use or capabilities. Applied to IT, such ideas have been pushed by the augmented reality domain where users can access a personalized view of the real world that helps them to carry out their activities. Although this idea has already been implemented in some ad hoc solutions (to exchange ratings for instance), we aim to provide a more generic solution. Our solution must be applicable to nowadays devices and applications with little adjustment to the underlying architectures. It should then be flexible enough to deal with the lack of standards in the domain without imposing architectural choices. Such lack of standard is very common in IT and mainly due to well known factors : (1) for technical reasons, developers tends to think that their âstandardâ are better suited for their current use-case, or/and (2) for commercial reasons companies want to keep a closed siloed system to capture their users, or/and (3) because the domain is still new and evolving and no standard as emerged yet, or/and finally (4) because the problem is too complex to be standardized and most proposed standards tend to be bloated and hard to use.

We are currently implementing these ideas by designing and experimenting two architectures/prototypes:

- **Matriona** is a global distributed framework developed on top of OSGi. This project has been described in more details in the previous activity report. It is meant to be a comprehensive framework for exposing devices as REST-like resources. Resources functionalities can be extended through the mean of decorators. The system also provides access control mechanisms. The main interest of matriona concerning the information enrichment is its ability to support dynamic extension of resource meta-information by application and to provide means to share these meta-information with others. It implements the concept of group of interest with access control on meta-information. The concepts described in Matriona are in the process to be published.
- **Little Thumb Registry (LTReg)** is an independent resource registry that provides the same enrichment capabilities than Matriona but impose less constraints on the architecture of application. Although the prototype is operational, Matriona does not comply with the principle advocated herebefore: it supposes the use of a pivot technology (REST) and assumes that application developers will develop their application on top of on OSGi based platform. The idea behind LTReg is to decouple the registry from the framework and to provide a registry in the manner of Consul [0] with meta-information enrichment and sharing mechanisms. By focussing only on the discovery mechanism and information sharing, LTReg imposes fewer constraints on application and comply more with the goal of being ready to use in actual application. This is still a work in progress.

6.3. Modeling activities to promote self-consumption of locally produced energy

[0] https://www.consul.io/

Participants: Jean-Marie Bonnin, Alexandre Rio, Yoann Maurel [contact].

Traditional electricity distribution schemes decouple the production sources from the consumers so that it is necessary to transport energy over long distances. This type of organization is illustrated by the consumption of region such as Brittany, where 91% of the energy consumed is imported. It induces inherent inefficiencies due to the line losses and the transformation steps and therefore induces a high infrastructure and distribution cost. To face these problems and in order to reduce the environmental impacts associated with the use of energy, recent years have seen the development of initiatives to produce energy locally.

The sources of renewable energies are good candidates for this because they are varied and adapt easily to the different geographical situations. The infrastructures necessary for their implementation also impose fewer constraints in terms of installation and safety. One of the main obstacles to the unique use of these technologies comes from their strong dependence on physical and meteorological characteristics, which makes it more difficult to foresee production capacities. These characteristics vary from one facility to another and from one region to another. The combined use of these technologies therefore appears to be necessary to ensure that there will always be available energy at the lowest possible cost. In this context, OKWind proposes to deploy self-production units directly where the consumption is done and has developed expertise in multi-source energy production (see section 8.1).

In 2016, we started to study a solution favoring maximum autonomy of the instrumented sites from the traditional channels energy production by modeling business processes and using learning algorithms to shift demanding activities according to local production capacities. For example, the system should be able to anticipate a potential consumption of hot water (and thus of the energy needed for its production) in order to produce it at the best time when the renewable energy is available. It should also choose the best storage solution for this energy: hot water could be directly stored by the heat pump for instance. The system must implement policies that will intelligently shift demanding activities according to the predictions of energy production. It thus requires:

- **capabilities to predict the production of energy**. A lot of theoretical work has been done in the literature to predict the production of renewable sources of energy. In addition, in order to evaluate the production of energy and its consumption over time, OKWind has developed data retrieval mechanisms on each deployed sites. They produce accurate statistics on production and consumption. Both approaches should be used as inputs of our decision processes and model. One of our goals is to evaluate the precision of the theoretical prediction models against these real-world data to determine which are the most relevant for the implementation of our approach.
- **capabilities to model the consumption on energy**. Numerous works of the literature are interested in similar problems but focuses mainly on building electricity consumption model of machine tools [10]. We propose to focus instead on activity and business processes. In a related domain, modeling work has been conducted on water consumption of farms [7]. The objective was to predict the water consumption of an operating farm by modeling business processes. Our goal is to propose a similar model for electricity targeting a broader scope of economic sectors.
- **capabilities to schedule activities in order to match production and consumption so as to promote self-consumption.** This requires developing control loop that will proactively analyze and predict consumption and take measure to shift demand. This can, in a first approach, be done by assisting the consumers and providing them guidance on when to perform certain tasks. Assisted demand shifting have already been developed for the residential domain [6] but this project focused on uses mainly and little on the modeling of business processes. Ultimately, we would like to develop automated process transparently when possible. The learning algorithms will be developed in collaboration with Ubiant [0], a company specialized in artificial intelligence to smart-buildings.

To validate the approach and to understand business processes, we have started a field study targeting two types of activities (e.g. farm or hotel). We also want to develop tools to simulate a site so that we can quickly evaluate our policies over simulated long periods of time.

[0] https://www.ubiant.com/en/about/

6.4. Definition of a Smart Energy Aware architecture

Participant: Jean-Marie Bonnin [contact].

In the past years, energy demand has increased and shifted especially towards electricity as the form of consuming energy. As the number of electric devices constantly grows and energy production must increasingly rely on renewable sources, this leads into noteworthy disparity between electricity production and consumption. Within the ITEA2 12004 Smart Energy Aware Systems (SEAS) project (see section 1), we proposed the "SEAS Reference Architecture Model" (S-RAM). This architecture relies on four distributed services that enable to interconnect any energy actors and give them the opportunity to provide new energy services. The benefits of S-RAM have been studied on a specific use case, which aims to provide a service for estimating local photovoltaic production. It particularly helps energy management systems better plan electric consumption. The main principles of this architecture have been published and we developed several profs of concept that have been demonstrated in the project consortium. Our partner continue to develop the components of the architecture that will be demonstrated in the final review of the project.

6.5. Context modeling for Smart Spaces

Participants: Yoann Maurel, Frédéric Weis [contact].

To provide services for Smart Building, 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 behaviors. Our aim is to compute small pieces of context we called *context attributes*. Those context attributes are diverse, for example a presence in a room, the number of people in a room etc. Building efficient and accurate context information using inexpensive and non-invasive sensors was and is still a great challenge. We proved, through the use of dedicated algorithms and a layered architecture that it is achievable when the targeted space (controlled environment) is known - due to the specific and non automated calibration process we used. Among all the available theories, we used the Belief Function Theory (BFT) [8] [9] as it allows to express **uncertainty** and **imprecision**.

Context is computed by a chain if three a tasks:

- 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. A belief function represents the degree of belief associated to each possible value of the context attribute.
- Then a set of belief functions (corresponding to a set of sensors) can be combined (fused).
- Finally the system can decide what is the "best" value for the context attribute.

Currently the BFT theories requires a huge calibration process. In 2016, we obtained new results on the semi-automated building of belief functions, that have to be provided by each sensor, using our BFT Java implementation (see section 5.1).

6.6. Towards Metamorphic Housing: the on-demand room

Participants: Frédéric Weis, Michele Dominici [contact].

6.6.1. A concrete example of Metamorphic Housing: the on-demand room

The research activities related to the research program on Metamorphic Housing mainly focused on defining the detailed architecture and functionalities of the selected case study, the on-demand room. We conducted an iterative co-design process, involving the partners of the chair "Habitat Intelligent et Innovation". Valuable input was also obtained by collaborating with Delta Dore, LOUSTIC, Université de Bretagne Occidentale, etc. The result was the identification of the needs of end users, building owners and managers with respect to the on-demand room. To satisfy these requirements, we proposed a system architecture, combining computer and mobile applications with domotic equipments and novel interaction means for end users.

These are inspired by the Pervasive Computing and Interactive Architecture principles, where a continuous and implicit interaction between occupants and the physical world is made possible by augmented architectural structures, which sense the natural actions of people and respond accordingly. In this way, the occupants of the dwellings equipped with on-demand room experience a new form of housing, stimulating social interactions between neighbors and satisfying periodic needs of additional housing surface, as we illustrated in [4]. We submitted our system architecture, novel interaction means and augmented structure designs to the industrial property services of Inria and University of Rennes 1, which are currently evaluating the possibility of establishing patent protection on these inventions.

6.6.2. Experimentation of Metamorphic Housing on social housing

We helped Néotoa, a social landlord, preparing and initiating an experimentation of the on-demand room on one of their residential buildings. For this, we built and coordinated a consortium of partners working on the project: Veolia, CCI Rennes, Cardinal Edifice, Rennes Métropole, Néotoa, Delta Dore, LOUSTIC, Université de Bretagne Occidentale, MobBI platform (University of Rennes 1), Inria, Institut de Gestion de Rennes. We took a user-centered approach to the problem, studying it from several points of view and mobilizing several disciplines: psychology and ergonomics (LOUSTIC), sociology (Université de Bretagne Occidentale), marketing (Institut de Gestion de Rennes). We conducted user interviews, initially leveraging the demonstrator of the on-demand room that we previously built via the Immersia virtual reality platform. Then, we ran on-line inquiries to reach a larger audience. We took into account the lessons that we learned in the design and development of a computing and domotic system, leveraging the expertise of valuable partners (Delta Dore, MobBI platform, Inria), as detailed in section 5.3.

7. Bilateral Contracts and Grants with Industry

7.1. SIMHet

Partner: YoGoKo

Starting: Nov 2015; ending: Oct 2018

Abstract: The SIMHet project is performed in partnership with YoGoKo, a start-up that develops innovative communication solutions for cooperative intelligent transport systems. The SIMHet project aims to develop a decision making mechanism that would be integrated in the ISO/ETSI ITS communication architecture. It will allow mobile devices or mobile routers to choose the best network interface for each embedded application/flow. For example, in a vehicular environment this mechanism could manage global (Internet) and local connections for each on board device/application, in order to ensure that applications and services are always best connected. Aware that "best" concept is context-dependent, such a decision making mechanism should take into account requirements from different actors (e.g., applications, user, network administrators) and contextual information. One of the difficulties is to take advantage of the knowledge the system could have about near future connectivity. In the vehicular context such information about the movement and the availability of network resources is available. If taking into account the future makes the decision making more complex, this could allow a better usage of network resources when they are available. Once current solutions in the market are based on very simple decisions, this smart mechanism will give competitive advantage for YoGoKo over its competitors.

8. Partnerships and Cooperations

8.1. Regional Initiatives

Project: Modélisation des activités de site consommateur d'énergie pour favoriser l'autoconsommation d'énergies renouvelables produites localement

Partner: OKWIND

Starting: Nov 2016; ending: Nov 2019

Abstract: OKWind [0] is a company specialized in local production of renewable energy. This project, with Inria DiverSE and TACOMA teams, aims at building a system that optimizes the use of different sources of renewable energy, choosing the most suitable source for the current demand and anticipating future needs, so as to favor the consumption of locally produced electricity. The system must be able to model clients' activities. It must also trigger actions (local consumption vs. local storage). The final goal is to use " locally produced" energy in a smarter way and to tend towards a self-consumption optimum.

Project: EkoHub

Partners: Ekolis, Delaye transport

Starting: Nov 2014; ending: Nov 2017

Abtract: The EkoHub project has been architectured around hors multi-technologies gateway and leverages on the one developped in the ITSSv6 European project. In addition to the multiple interfaces of our platforms, sensor devices have been incorporated into the project and we studied different scenarios elaborated with our professional partners (Layaye Logistics). Intelligent data management schemes are being studied to adapt to the communication environment and the needs of the application consuming the data.

8.2. National Initiatives

Project: Pervasive_RFID

Partner: IETR

Starting: July 2013; ending: July 2016

Abtract: Pervasive_RFID is a joint effort (within the CominLabs initiative, see http://www.cominlabs.ueb.eu/) started in July 2013 with IETR (institut d'électronique et de télécommunications de Rennes) to study and design innovative RFID reading protocols in the context of pervasive computing applications. Some limitations of existing RFID technology become challenging: unlike standard RFID application scenarios, pervasive computing often involves uncontrolled environment for RFID, where tags and reader have to operate in much more difficult situations that those usually encountered or expected for classical RFID systems.

Project: GLIE - Guidage Lumineux par l'Intelligence de l'Environnement

Partner: OyaLight

Starting: December 2014; ending: April 2016

Abstract: GLIE is a collaborative projet with OYALIGT and TACOMA group. The objective of the project is to design and demonstrate a new service combining connected LEDs provided by OYALIGHT and a software tool developed by TACOMA. By integrating and analyzing data transmitted by the sensors integrated into LEDs, the service must be able to detect a given context and to react accordingly.

Project: Greenfeed

http://greenfeed.org

Partner: BeNomad, Mines St Etiennes, Enedis, G2MOBILITY, GreedPocket

Starting: July 2013; ending: Nov 2016

[0] http://www.okwind.fr/

Abstract: Greenfeed aims at improving electro-mobility, which means the ease with which users can travel using electric cars. In order to achieve its goal, the project focuses three main operators: electro-mobility service provider (EMSP), distribution service operator (DSO), and charging station operator (CSO). During the project, the role of these actors have been precisely defined, so were the role of the systems they were in charge of. A great effort has been put on interoperability, so that the developed systems could collaborate with each other. One of the key use case was to enable the smart management of available power on a 10 charging site. This led to a demonstration in which a Renault Zoé, customized by the Institut Védécom, was able to negotiate a charge planning with the electric power grid. Then a grid initiated renegotiation was demonstrated, once the initial smart charging process had began. This was the first time this behavior had been achieved with a vehicle in France.

8.3. European Initiatives

8.3.1. FP7 & H2020 Projects

Project acronym: SCOOP@F part 2

Partners: MEDE, Renault, PSA

Starting: January 2016; ending: Dec 2018

Abstract: SCOOP@F is a Cooperative ITS pilot deployment project that intends to connect approximately 3000 vehicles with 2000 kilometers of roads. It consists of 5 specific sites with different types of roads: Ile-de-France, "East Corridor" between Paris and Strasbourg, Brittany, Bordeaux and Isère. SCOOP@F is composed of SCOOP@F Part 1 from 2014 to 2015 (ongoing) and SCOOP@F Part 2 from 2016 to 2018. Its main objective is to improve the safety of road transport and of road operating staff during road works or maintenance. The project includes the validations of Cooperative ITS services in open roads, cross border tests with other EU Member States (Spain, Portugal and Austria) and development of a hybrid communication solution (3G-4G/ITS G5).

8.3.2. Collaborations in European Programs, Except FP7 & H2020

Project acronym: SEAS (ITEA3)

Partners: Telecom Paris Tech, Telecom Saint Etienne, Mines Saint Etienne, Engie, Kerlink, BeNomad, ICAM, CNR, VTT

Starting: Feb 2014; ending: Jan 2017

Abstract: The SEAS project addresses the problem of inefficient and unsustainable energy consumption, which is due to a lack of sufficient means to control, monitor, estimate and adapt the energy use of systems versus the dynamic use situations and circumstances influencing the energy use. The objective of the SEAS project is to enable energy, ICT and automation systems to collaborate at consumption sites, and to introduce dynamic and refined ICT-based solutions to control, monitor and estimate energy consumption. Proposed solution should enable energy market participants to incorporate micro-grid environments and active customers. We are involved in the project to design a distributed system architecture and to implement two proofs of concept: the first one is related to the electric vehicle charging and the other one to the prevision of solar energy production.

9. Dissemination

9.1. Promoting Scientific Activities

9.1.1. Scientific Events Organisation

9.1.1.1. Member of the Organizing Committees

- PC member for Ehpwas 2016: 4th IEEE Int. Workshop on E-Health , 2016, F. Weis
- PC member of The 83th IEEE Vehicular Technology Conference, 15-18 May 2016, Nanjing, China, JM. Bonnin
- PC member of EAI International Conference on Mobile Medical Multimedia Technologies, Applications and Services, 14-15 June 2016, Budapest, Hungary, JM. Bonnin
- PC member of 13th African Conference on Research in Computer Science and Applied Mathematics, 11 - 14 October 2016, Tunis, Tunisia, JM. Bonnin
- PC member of The 84th IEEE Vehicular Technology Conference, 18-21 September 2016, Montreal, Canada, JM. Bonnin

9.1.2. Scientific Events Selection

9.1.2.1. Reviewer

- Mobile Networks and Applications, JM. Bonnin
- IEEE transactions on Mobile Computing, JM. Bonnin
- Simulation Modeling Practice and Theory, JM. Bonnin
- Computer Standards & Interfaces, an international journal on Engineering Science and Technology, JM. Bonnin

9.1.3. Invited Talks

- Keynote speech at the 6th NGNS international conferences, 17/12/2016, Rabat, Morocco, JM. Bonnin
- Invited talk at the Drive for All seminary, 31/08/2016, Paris, France, JM. Bonnin
- RFIDay Event, June 2016, Vitré. Talk and demos, P. Couderc
- Invited talk at the BMW Summer School, "New Business, New Mobility - YoGoKo case study", July 2016, Tegernsee, Bavaria, Germany, JM. Bonnin
- SESAME school of Tunis, Nov 2016, Tunis, Tunisia, JM. Bonnin
- Invited Talk at Smart Systems inter Labex Symposium, November 2016, Besançon, P. Couderc
- ENSAM school of Meknes, Dec 2016, Meknes, Morocco, JM. Bonnin

9.1.4. Leadership within the Scientific Community

- Jean-Marie Bonnin is member of the scientific council of the GIS ITS
- Jean-Marie Bonnin is member of the scientific council the Id4Car cluster
- Jean-Marie Bonnin is an elected member of the Scientific Council of IMT

9.1.5. Scientific Expertise

- Evaluation committee for the Belgium government, JM. Bonnin
- Expert for ANR, F. Weis and JM Bonnin
- Expert for CSV board of "Pôle Images et Réseaux", P. Couderc

9.1.6. Research Administration

- Head of the Networks, Telecommunication and service department at IRISA, JM. Bonnin

9.2. Teaching - Supervision - Juries

9.2.1. Teaching

L2/L3: network computing (lectures, tutorials, labs), 250 hours, F. Weis, University Rennes 1

Master: Wireless LANs, F. Weis, 20 hours, M2, Telecom Bretagne

Master: Supervision of a Master 1 project related to the smart building (in collaboration with Myriads), 48 hours, P. Couderc, University Rennes 1

Master 2: Pervasive computing introduction course, Paul Couderc, 4 hours, University Rennes 1

Master 1: Network programming (lectures, tutorial, labs), 78 hours, Y. Maurel

L3/M2: network communications protocol for building automation (lectures, labs), 80 hours, Y. Maurel

Master 2: Software engineering (lectures, tutorial, labs), 82 hours, Y. Maurel

9.2.2. Supervision

PhD in progress: Adrien Capaine, Vers une plate-forme de LED connectées comme vecteur de services contextuels dans le cadre des bâtiments intelligents, 01/05/15, Frédéric Weis and Yoann Maurel

PhD in progress: Zaineb Lioune, Une Architecture pour des Services e-Santé évolutifs dans le cadre des Maisons Intelligentes, 01/09/14, Frédéric Weis, Tayeb Lemlouna and Philippe Roose

PhD in progress: Indra Ngurah, Protocoles de routage pour l'ITS, 01/05/16, Jean-Marie Bonnin

PhD in progress: Christophe Couturier, Exploitation de mécanismes EDGE pour l'ITS, 01/11/16, Jean-Marie Bonnin

PhD in progress: Alexandre Rio, Modélisation des activités de site consommateur d'énergie, 01/10/16, Olivier Barais and Yoann Maurel

9.2.3. Juries

Clément DUHART, "Toward Organic Ambient Intelligences ? EMMA", University du Havre, F. Weis, PhD referee

Minh Huong NGUYEN, "Contribution to Intelligent Transportations Systems: Security of Safety applications in Vehicular Ad hoc Networks", JM Bonnin, PhD referee

Matthieu KANJ, "Intelligent Supervision of Flexible Optical Networks", JM Bonnin, Président

9.3. Popularization

- Invited talk at the Telecom ParisTalks on "De la voiture connectée à la voiture autonome", "Le véhicule autonome, un véhicule ultra-communicant", March 2016, Paris, France, JM. Bonnin
- Invited talk at the A3C7 public event: "L?automobile intelligente et communicante: Carminat, GPS et demain?", "L?automobile communicante dans la ville", June 2016, Cesson-Sévigné, France, JM. Bonnin
- Invited talk at the workshop "Atelier Cybercriminalité dans les transports - Véhicule autonome et transport connecté : quelles stratégies de fiabilité et de sécurité ?","Sécurité des Communications pour les ITS", June 2016, Angers, France, JM. Bonnin
- "Les véhicules autonomes seront communicants?, Revue TELECOM 182, Oct 2016, JM. Bonnin
- Special issue TPC for Revue TEC mobilité intelligente : "Routes & Véhicules : Quelles Connexions ?", Nov 2016, JM. Bonnin
- "Standards et ITS : vers un écosystème ouvert ?", Revue TEC Mobilité intelligente, Nov 2016, JM. Bonnin
- "Quand les modes de vie actuels façonnent l'habitat du futur", Site Internet Batiactu, Jan 2016, http://www.batiactu.com/edito/quand-modes-vie-actuels-faconnent-habitat-futur-43438.php
- Invited talk "Le logement évolutif, Michele Dominici", Matin de l'aura angevine, Jan 2016, Angers
- Cardinal Edifice imagine l'habitat de demain, Communiqué de presse de Cardinal, relayé sur le site Batinfo, Jan 2016. http://batinfo.com/actualite/cardinal-edifice-imagine-lhabitat-de-demain_4339

- Invited talk "Ville et Habitat intelligent : des innovations au service du vivre-ensemble ?", Michele Dominici, Feb 2016, La Friche la Belle de Mai, Marseille
- "Place au logement caméléon - freemium" - Métiers. LeMoniteur, 1 mars 2016. http://www.lemoniteur.fr/articles/place-au-logement-cameleon-31516795
- Invited talk "L'habitat d'hier, d'aujourd'hui et de demain, intervention de Michele Dominici", Apr 2016 dans le cadre du Printemps du Développement Durable, Langueux
- Ecole d'été "Ecologie et biens communs", Université de Bretagne Occidentale, Sept 2016, Michele Dominici

10. Bibliography

Publications of the year

Doctoral Dissertations and Habilitation Theses

[1] F. J. ACOSTA PADILLA.*Self-adaptation for Internet of Things applications*, Université de Rennes 1, France, December 2016, https://hal.inria.fr/tel-01426219.

International Conferences with Proceedings

[2] Z. LIOUANE, T. LEMLOUMA, P. ROOSE, F. WEIS, M. HASSANI.*A Genetic–based Localization Algorithm for Elderly People in Smart Cities*, in "The 19th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems", Malte, Malta, November 2016, https://hal.inria.fr/hal-01372772.

[3] Z. LIOUANE, T. LEMLOUMA, P. ROOSE, F. WEIS, M. HASSANI.*A Markovian-based Approach for Daily Living Activities Recognition*, in "The International Conference on Sensor Networks (SENSORNETS'16)", rome, Italy, February 2016, https://hal.inria.fr/hal-01280001.

Research Reports

[4] M. DOMINICI, M. BANÂTRE.*Virtualizing the physical world: the on-demand room case study*, Irisa, October 2016, n° 2039, https://hal.inria.fr/hal-01417204.

References in notes

[5] R. BALDONI, M. BANÂTRE, M. DOMINICI, D. DUMORTIER, M. EL MANKIBI, M. FRÉJUS, B. KOLDEHOFE, M. MECELLA, K. ROTHERMEL, P. SUIGNARD.*Comfort-Oriented Metamorphic House (COMETAE)*, October 2012, n° PI 1998, Publications Internes de l'IRISA ISSN : 2102-6327, https://hal.inria.fr/hal-00864884.

[6] J. BOURGEOIS.*Interactive Demand-Shifting in the Context of Domestic Micro-Generation*, The Open University ; Université de Rennes 1 [UR1], June 2016, https://hal.inria.fr/tel-01385022.

[7] J.-M. BRUEL, B. COMBEMALE, I. OBER, H. RAYNAL.*MDE in Practice for computational science*, in "Procedia Computer Science", 2015, vol. 51, p. 660–669.

[8] B. PIETROPAOLI, M. DOMINICI, F. WEIS.*Propagation of Belief Functions through Frames of Discernment : Application to Context Computing*, in "The 26th International FLAIRS Conference", St. Pete Beach, United States, May 2013, https://hal.inria.fr/hal-00927088.

[9] B. PIETROPAOLI, M. DOMINICI, F. WEIS.*Virtual Sensors and Data Fusion in a Multi-Level Context Computing Architecture*, in "16th International Conference on Information Fusion (FUSION 2013)", Instanbul, Turkey, July 2013, https://hal.inria.fr/hal-00927092.

[10] L. ZHOU, J. LI, F. LI, Q. MENG, J. LI, X. XU.*Energy consumption model and energy efficiency of machine tools: a comprehensive literature review*, in "Journal of Cleaner Production", 2016, vol. 112, p. 3721–3734.

Team TAMIS

Threat Analysis and Mitigation for Information Security

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

RESEARCH CENTER
Rennes - Bretagne-Atlantique

THEME
Security and Confidentiality

Table of contents

Team TAMIS

Creation of the Team: 2016 January 01

Keywords:

Computer Science and Digital Science:

4. - Security and privacy
4.1. - Threat analysis
4.2. - Correcting codes
4.4. - Security of equipment and software
4.5. - Formal methods for security
4.8. - Privacy-enhancing technologies
4.9. - Security supervision

Other Research Topics and Application Domains:

6. - IT and telecom
6.4. - Internet of things
6.5. - Information systems
6.6. - Embedded systems
8.1. - Smart building/home
8.2. - Connected city
8.4. - Security and personal assistance
9.8. - Privacy
9.9. - Risk management
9.10. - Ethics

1. Members

Research Scientists

Axel Legay [Team leader, Inria, Researcher, HDR]
Christian Grothoff [Inria, Advanced Research position]
Annelie Heuser [CNRS, Researcher, from Oct 2016]
Olivier Zendra [Inria, Researcher, from Sep 2016]

Faculty Members

Fabrizio Biondi [Supelec, Associate Professor Inria "Chaire Malware" since Nov 2016; was Inria Engineer before]
Jean-Louis Lanet [Univ. of Limoges, Professor, HDR]
Kim Larsen [Univ. of Aalborg, Professor, Inria "Chaire internationale", HDR]

Technical Staff

Jeffrey Paul Burdges [Inria]
Francois-Renaud Escriva [DGA]
Ulrich Fahrenberg [Ecole Polytechnique, until Sep 2016]
Thomas Given-Wilson [Inria]
Sebastien Josse [DGA]
Colas Le Guernic [DGA]
Laurent Morin [Inria]

Van-Chan Ngo [Inria, until Feb 2016]
Jean Quilbeuf [Inria]
Sean Sedwards [Inria]
Marcello Stanisci [Inria]
Gabor Toth [Inria]
Louis-Marie Traonouez [Inria]

PhD Students

Sebanjila Bukasa [Inria]
Mounir Chadli [Algerian Embassy in France]
Olivier Decourbe [Inria, from Mar 2016]
Florian Dold [Inria]
Mihai Enescu [Inria, from Oct 2016]
Alvaro Garcia Recuero [Inria, until Oct 2016]
Alexandre Gonzalvez [Telecom Bretagne, from Apr 2016]
Nisrine Jafri [Inria]
Rokia Lamrani Alaoui [Inria, until Jun 2016]
Tristan Ninet [Thales, from Dec 2016, granted by CIFRE]
Aurelien Palisse [Inria]

Post-Doctoral Fellows

Ronan Lashermes [Inria, from Mar 2016]
Hélène Le Bouder [Inria]

Visiting Scientists

Razika Lounas [University of Boumerdès (Algeria), until Jun 2016]
Abdelhak Mesbah [University of Boumerdès (Algeria), Visiting PhD Student, from Apr 2016 until May 2016]
Noreddine El Janati El Idrissi [University of Rabat (Morrocco), Visiting PhD Student, from May 2016 until Jun 2016]

Administrative Assistant

Cecile Bouton [Inria]

Others

Antoine Durand [University of Bordeaux, Master 1 Student, from Jun 2016 until Sep 2016]
Gregory Durand [University of Limoges, Master Student, from Mar 2016 until Aug 2016]
Christophe Genevey-Metat [University of Rennes 1, Master Student , from May 2016 until Jul 2016]
Marouan Sami [University of Limoges, Master 1 Student, from Jul 2016 until Aug 2016]

2. Overall Objectives

2.1. Context

Security devices are subject to drastic security requirements and certification processes. They must be protected against potentially complex exploits that result from the combination of software and hardware attacks. As a result, a major effort is needed to develop new research techniques and approaches to characterize security issues, as well as to discover multi-layered security vulnerabilities in complex systems.

In recent years, we have witnessed two main lines of research to achieve this objective.

The first approach, often called *offensive security*, relies on engineering techniques and consists in attacking the system with our knowledge on its design and our past expertise. This is a creative approach that supports (1) checking whether a system is subject to existing vulnerabilities, i.e. classes of vulnerabilities that we already discovered on other systems, and (2) discovering new types of vulnerabilities that were not foreseen and that may depend on new technologies and/or programming paradigms. Unfortunately, this approach is limited to systems whose complexity remains manageable at the human level. This means that exploits which combine several vulnerabilities may be hard to identify. The second and more formal approach builds on formal models (also known as *formal methods*) to automatically detect vulnerabilities, or prove their absence. This is applicable to systems whose complexity is beyond human reasoning, but can only detect existing classes of vulnerabilities, i.e., those that have been previously characterized by offensive security.

2.2. Approach and motivation

The claim made by TAMIS is that *assessing security requires combining both engineering and formal techniques*.

As an example, security exploits may require combining classes of well-known vulnerabilities. The detection of such vulnerabilities can be made via formal approaches, but their successful combination requires human creativity. TAMIS's central goal is thus to demonstrably narrow the gap between the vulnerabilities found using formal verification and the issues found using systems engineering. As a second example, we point out that there are classes of attacks that exploit both the software and hardware parts of a system. Although vulnerabilities can be detected via formal methods in the software part, the impact of attacking the hardware still needs to be modeled. This is often done by observing the effect of parameter changes on the system, and capturing a model of them. To address this situation, the TAMIS team bundled resources from scalable formal verification and secure software engineering for *vulnerability analysis*, which we extend to provide methods and tools to (a) *analyze (binary) code including obfuscated malware*, and (b) *build secure systems*.

Very concrete examples better illustrate the differences and complementarity of engineering and formal techniques. First, it is well-known that formal methods can be used to detect buffer overflows. However, the definition of buffer overflows itself was made first in 1972 when the Computer Security Technology Planning study laid out the technique and claimed that over sizing could be exploited to corrupt a system. This exploit was then popularized in 1988 as one of the exploits used by the Morris worm, and only at that point systematic techniques were developed to detect it. Another example is the work we conducted in attacking smart cards. The very firsts experiments were done at the engineering level, and consisted of retrieving the key of the card in a brute force manner. Based on this knowledge, we generated user test-cases that characterize what should not happen. Later, those were used in a fully automatized model-based testing approach [71].

3. Research Program

3.1. Axis 1: Vulnerability analysis

This axis proposes different techniques to discover vulnerabilities in systems. The outcomes of this axis are (a) new techniques to discover system vulnerabilities as well as to analyze them, and (b) to understand the importance of the hardware support.

Most existing approaches used at the engineering level rely on testing and fuzzing. Such techniques consist in simulating the system for various input values, and then checking that the result conforms to a given standard. The problem being the large set of inputs to be potentially tested. Existing solutions propose to extract significant sets by mutating a finite set of inputs. Other solutions, especially concolic testing developed at Microsoft, propose to exploit symbolic executions to extract constraints on new values. We build on those existing work, and extend them with recent techniques based on dissimilarity distances and learning. We also account for the execution environment, and study techniques based on the combination of timing attacks with fuzzing techniques to discover and classify classes of behavior of the system under test.

Techniques such as model checking and static analysis have been used for verifying several types of requirements such as safety and reliability. Recently, several works have attempted to adapt model checking to the detection of security issues. It has clearly been identified that this required to work at the level of binary code. Applying formal techniques to such code requires the development of disassembly techniques to obtain a semantically well-defined model. One of the biggest issues faced with formal analysis is the state space explosion problem. This problem is amplified in our context as representations of data (such as stack content) definitively blow up the state space. We propose to use statistical model checking (SMC) of rare events to efficiently identify problematic behaviors.

We also seek to understand vulnerabilities at the architecture and hardware levels. Particularly, we evaluate vulnerabilities of the interfaces and how an adversary could use them to get access to core assets in the system. One particular mechanism to be investigated is the DMA and the so-called Trustzone. An ad-hoc technique to defend against adversarial DMA-access to memory is to keep key material exclusively in registers. This implies co-analyzing machine code and an accurate hardware model.

3.2. Axis 2: Malware analysis

Axis 1 is concerned with vulnerabilities. Such vulnerabilities can be exploited by an attacker in order to introduce malicious behaviors in a system. Another method to identify vulnerabilities is to analyze malware that exploits them. However, modern malware has a wide variety of analysis avoidance techniques. In particular, attackers obfuscate the code leading to a security exploit. For doing so, recent black hat research suggests hiding constants in program choices via polynomials. Such techniques hinder forensic analysis by making detailed analysis labor intensive and time consuming. The objective of research axis 2 is to obtain a full tool chain for malware analysis starting from (a) the observability of the malware via deobfuscation, and (b) the analysis of the resulting binary file. A complementary objective is to understand how hardware attacks can be exploited by malwares.

We first investigate obfuscation techniques. Several solutions exist to mitigate the packer problem. As an example, we try to reverse the packer and remove the environment evaluation in such a way that it performs the same actions and outputs the resulting binary for further analysis. There is a wide range of techniques to obfuscate malware, which includes flattening and virtualization. We will produce a taxonomy of both techniques and tools. We will first give a particular focus to control flow obfuscation via mixed Boolean algebra, which is highly deployed for malware obfuscation. We recently showed that a subset of them can be broken via SAT-solving and synthesis. Then, we will expand our research to other obfuscation techniques.

Once the malware code has been unpacked/deobfuscated, the resulting binary still needs to be fully understood. Advanced malware often contains multiple stages, multiple exploits and may unpack additional features based on its environment. Ensuring that one understands all interesting execution paths of a malware sample is related to enumerating all of the possible execution paths when checking a system for vulnerabilities. The main difference is that in one case we are interested in finding vulnerabilities and in the other in finding exploitative behavior that may mutate. Still, some of the techniques of Axis 1 can be helpful in analyzing malware. The main challenge for axis 2 is thus to adapt the tools and techniques to deal with binary programs as inputs, as well as the logic used to specify malware behavior, including behavior with potentially rare occurrences. Another challenge is to take mutation into account, which we plan to do by exploiting mining algorithms.

Most recent attacks against hardware are based on fault injection which dynamically modifies the semantics of the code. We demonstrated the possibility to obfuscate code using constraint solver in such a way that the code becomes intentionally hostile while hit by a laser beam. This new form of obfuscation opens a new challenge for secure devices where malicious programs can be designed and uploaded that defeat comprehensive static analysis tools or code reviews, due to their multi-semantic nature. We have shown on several products that such an attack cannot be mitigated with the current defenses embedded in Java cards. In this research, we first aim at extending the work on fault injection, then at developing new techniques to analyze such hostile code. This is done by proposing formal models of fault injection, and then reusing results from our work on obfuscation/deobfuscation.

3.3. Axis 3: Building a secure network stack

To evaluate the techniques developed in Axes 1 and 2, we analyze concrete systems developed not only with industry partners, but also within the team. By using our own systems, we can co-evolve best-practices, while externally developed systems provide realistic challenges especially with respect to analyzing obfuscated malware in the hardware or complex vulnerabilities. In this context, Christian Grothoff (ARP Inria) is currently developing a new Internet, which is supposed to be more secure. This introduces interesting challenges both in terms of vulnerability and malware analysis, and hence should be a great opportunity to mix the competences of all the members of the team.

More precisely, this system intends to challenge the idea that network security is an administrative task, where network administrators shield users with passwords, firewalls, intrusion detection systems and policies. Instead, we want to eliminate administrators that have power over user's data, and as such administrators themselves are liabilities, and because a network design that permits administrative intrusion inherently adds vulnerabilities. Instead, the system should ensure secure communication mechanisms without trusted third parties.

Key challenges we work on include (a) improving scalable secure ad-hoc decentralized routing, including key-value lookup, unicast and multicast communication, (b) protecting meta-data in the overlay using advanced decentralized onion routing, (c) a unified public-key infrastructure and identity management solution that is suitable to replace the Web-of-Trust, X.509, DNSSEC and other legacy methods for naming and identifying services, (d) secure synchronous and asynchronous messaging at scale, providing decentralized alternatives to common online social applications and addressing challenges in protocol evolution and compatibility. Finally, we are currently working on GNU Taler, a new secure privacy-preserving payment system where users never have to authenticate. This system in particular can be used as a concrete test case for the methods developed in the team.

To support this research work, we develop a framework named GNUnet. It provides a clear separation into layers, which facilitates testing and verifying the various components. However, we see that often existing formal verification techniques still do not scale to typical subsystems encountered in practice. Our objective is thus to exploit efficient and scalable formal techniques techniques proposed in Axis 1 together with engineering skills in order to guide the validation (message synchronization, data protection, ...) and reach the best compromise. An additional complication is that we need a validation process that not merely covers the software itself, but also all of its dependencies (such as database, cryptographic libraries and networking libraries). For the Taler-specific hardware, we are envisioning an NFC-powered device, which creates new challenges in terms of securing cryptographic computations in a setting where the adversary has control over the power supply. In such a case, the attacker can drive the environment and modify the behavior of the system as we have shown in Axis 2. Providing the control of the environment is a new vector for attackers.

4. Application Domains

4.1. System analysis

The work performed in Axes 1 and 2 and the methods developed there are applicable to the domain of system analysis, both wrt. program analysis and hardware analysis.

4.2. Cybersecurity

The work done in the 3 axes above aims at improving cybersecurity, be it via vulnerability analyses, malware analyses and the development of safer networking mechanisms.

4.3. Safe Internet

The work done in Axis 3 above very directly contributes to the goal of a safer Internet.

5. Highlights of the Year

5.1. Highlights of the Year

New major release of Plasma Lab

Participants: Axel Legay, Sean Sedwards, Louis-Marie Traonouez.

We have released version 1.4.0 of our Plasma Lab software. This new version introduces a new command line interface for launching Plasma Lab. Besides the Graphical Interface, most of Plasma Lab functionalities are now available directly from the command line. Additionally the new version includes a new algorithm for cross entropy minimization using importance sampling. It allows to estimate the probabilities of rare events.

Fault injection proof-of-concept

Participants: Axel Legay, Jean-Louis Lanet, Thomas Given-Wilson, Nisrine Jafri.

Creation of a proof of concept to show that formal verification can be used to discover fault injections induced by hardware attacks.

Creation of LHS platform

Participants: Jean-Louis Lanet, Hélène Le Bouder, Ronan Lashermes.

Entry into service of the LHS platform that can be used to monitor systems, inject faults, or reason on ransomware.

Taler Systems startup creation

Participants: Jeffrey Burdges, Florian Dold, Christian Grothoff, Marcello Stanisci.

A startup, Taler Systems S.A. was formally created, and we started the contractual paperwork required. An interview was given to RWGV-Genossenschaftsblatt (an internal publication of a large group of German banks).

Contract with CISCO

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

Signature of a major research collaboration contract between Tamis and CISCO to work on malware analysis. The collaboration will fund 3 engineers, trips to visit CISCO and participate to conferences on the topic, as well as a powerful servers to store and analyse malware.

Awards

Axel Legay received the first Parnass award.

Christian Grothoff became an Ashoka fellow.

6. New Software and Platforms

6.1. MHD

GNU libmicrohttpd
KEYWORDS: Embedded - Web 2.0
SCIENTIFIC DESCRIPTION

We are providing a standards compliant and complete implementation of the HTTP server protocol that allows developers to easily write correct HTTP servers. Key challenges include code size minimization (for IoT devices), performance (zero copy, scalability to 100k concurrent connections), portability and security. MHD is already widely used in production by both academic and industrial users. Ongoing research challenges include formal verification.

FUNCTIONAL DESCRIPTION

GNU libmicrohttpd is a small C library that is supposed to make it easy to run an HTTP server as part of another application.

- Participants: Evgeny Grin, Christian Grothoff and Sree Hrsha Totakura
- Partner: The GNU Project
- Contact: Christian Grothoff
- URL: http://www.gnu.org/software/libmicrohttpd/

6.2. PLASMA Lab

KEYWORDS: Model Checking - Statistical - Model Checker - Runtime Analysis - Security - Code analysis - Statistics - Energy - Aeronautics

SCIENTIFIC DESCRIPTION

Statistical model checking (SMC) is a fast emerging technology for industrial scale verification and optimisation problems. Plasma was conceived to have high performance and be extensible, using a proprietary virtual machine. Since SMC requires only an executable semantics and is not constrained by decidability, we can easily implement different modelling languages and logics.

FUNCTIONAL DESCRIPTION

PLASMA Lab is a compact, efficient and flexible platform for statistical model checking of stochastic models. PLASMA Lab demonstrates the following advances: -Use your own simulator and checker via our plugin system. -Build your software around Plasma Lab using our API. -Prism (Reactives Modules Language-RML) and Biological languages supported. -Matlab and SytemC plugins. -Distributed architecture. Whether you plan to use several computers on a local area network or a grid, you can run PLASMA Lab in an easy way. -Fast algorithms. -Efficient data structure, low memory consumption. -Developed with Java for compatibility.

- Participants: Axel Legay, Sean Sedwards, Benoit Boyer, Louis-Marie Traonouez, Kevin Corre and Matthieu Simonin
- Contact: Axel Legay
- URL: http://plasma-lab.gforge.inria.fr/plasma_lab_doc/1.4.0/html/introduction.html#

6.3. Quail

FUNCTIONAL DESCRIPTION

Privacy is a central concern for Systems of Systems and interconnected objects. We propose QUAIL, a tool that can be used to quantify privacy of components. QUAIL is the only tool able to perform an arbitrary-precision quantitative analysis of the security of a system depending on private information. Thanks to its Markovian semantics model, QUAIL computes the correlation between the system's observable output and the private information, obtaining the amount of bits of the secret that the attacker will infer by observing the output.

- Participants: Fabrizio Biondi, Axel Legay, Louis-Marie Traonouez and Andrzej Wasowski
- Contact: Axel Legay
- URL: https://project.inria.fr/quail/

6.4. GNUnet

SCIENTIFIC DESCRIPTION

The GNUnet project seeks to answer the question what a modern Internet architecture should look like for a society that care about security and privacy. We are considering all layers of the existing well-known Internet, but are also providing new and higher-level abstractions (such as voting protocols, Byzantine consensus, etc.) that are today solved in application-specific ways. Research questions include the desired functionality of the overall stack, protocol design for the various layers as well as implementation considerations, i.e. how to implement the design securely.

FUNCTIONAL DESCRIPTION

GNUnet is a framework for secure peer-to-peer networking that does not use any centralized or otherwise trusted services. Our high-level goal is to provide a strong free software foundation for a global network that provides security and in particular respects privacy.

GNUnet started with an idea for anonymous censorship-resistant file-sharing, but has grown to incorporate other applications as well as many generic building blocks for secure networking applications. In particular, GNUnet now includes the GNU Name System, a privacy-preserving, decentralized public key infrastructure.

- Participants: Christian Grothoff, Florian Dold, Jeffrey Paul Burdges, Gabor Toth, Sree Hrsha Totakura and Alvaro Garcia Recuero
- Partner: The GNU Project
- Contact: Christian Grothoff
- URL: https://gnunet.org/

6.5. Taler

GNU Taler
KEYWORD: Privacy
SCIENTIFIC DESCRIPTION

Taler is a Chaum-style digital payment system that enables anonymous payments while ensuring that entities that receive payments are auditable. In Taler, customers can never defraud anyone, merchants can only fail to deliver the merchandise to the customer, and payment service providers can be fully audited. All parties receive cryptographic evidence for all transactions, still, each party only receives the minimum information required to execute transactions. Enforcement of honest behavior is timely, and is at least as strict as with legacy credit card payment systems that do not provide for privacy.

The key technical contribution underpinning Taler is a new refresh protocol which allows fractional payments and refunds while maintaining untraceability of the customer and unlinkability of transactions. The refresh protocol combines an efficient cut-and-choose mechanism with a link step to ensure that refreshing is not abused for transactional payments.

We argue that Taler provides a secure digital currency for modern liberal societies as it is a flexible, *libre* and efficient protocol and adequately balances the state's need for monetary control with the citizen's needs for private economic activity.
FUNCTIONAL DESCRIPTION

Taler is a new electronic payment system. It includes an electronic wallet for customers, a payment backend for merchants and the main payment service provider logic called the exchange. Taler offers Chaum-style anonymous payments for citizens, and income-transparency for taxability.

- Participants: Jeffrey Paul Burdges, Marcello Stanisci, Florian Dold, Gabor Toth and Christian Grothoff
- Partner: The GNU Project
- Contact: Christian Grothoff
- URL: http://taler.net/

6.6. VITRAIL - Visualisation Tool

Real-Time, Advanced, Immersive Visualization of Software / Visualizer
KEYWORD: Visualization of software
SCIENTIFIC DESCRIPTION

It is difficult for developers to explore and understand the source code of large programs, for example in objet-oriented languages programs featuring thousands of classes. Visualization methods based on daily life metaphors have thus been proposed. The VITRAIL Visualization tool (or VITRAIL Vizualizer) makes it possible to display, visualize and explore Java programs in a metaphorical way, using the city metaphor. An execution trace of the Java (byte)code provided by VITRAIL JBInstrace tool, is provided as input to VITRAIL Visualizer which displays a city-like metaphorical world showing the static structure of the code as well as some dynamic elements (calls).
This program may be used in Tamis as a basis for tools for the visualization of security events in programs.

FUNCTIONAL DESCRIPTION

This program makes it possible to displays, visualizes and explores Java programs in a metaphorical way (using the city metaphore). Useful for complex application developers/architects.

- Participants: Damien Bodenes, Olivier Zendra and Olivier Demengeon
- Contact: Olivier Zendra

6.7. VITRAIL 6 JBInsTrace

Real-Time, Advanced, Immersive Visualization of Software / Java Bytecode Instrumenter and Tracer
KEYWORDS: Java - Bytecode - Instrumentation - Profiling - Execution trace - Basic block
SCIENTIFIC DESCRIPTION

VITRAIL JBInsTrace is a program to instrument Java bytecode to trace its execution. The trace contains both static and dynamic information (calls). It is produced by intercepting the JVM class loader and replacing it by ours. Thus Java bytecode file are not modified, since instrumentation is performed on the fly, in memory. This makes it possible to instrument the whole program code, including libraries. Java source code is not needed. The trace which is then fed into our program VITRAIL Visualizer is an XML-like file.
This program may be used in Tamis as a basis for tools to instrument Java bytecode for security.

FUNCTIONAL DESCRIPTION

VITRAIL JBInsTrace is a program to instrument Java bytecode files to trace their execution. The trace is then fed into our VITRAIL Visualizer tool.

- Participants: Pierre Caserta and Olivier Zendra
- Contact: Olivier Zendra

6.8. Platforms

6.8.1. Malware'o'Matic

This LHS platform is dedicated to the collect, the categorization and the analyze of malware. We are currently interested in a specific kind of malware the ransomware. The platform grabs periodically samples of public data bases, executes the ransomware without virtualization on a victim PC and evaluate the implemented detection mechanisms. Once a ransomware has been executed the image of the OS is automatically restored and a new sample is evaluated. The platform is fully automatic and target Windows platforms (seven, W10) in both 32 bits and 64 bits versions.

6.8.2. Faustine

This LHS platform is dedicated to the EM fault injection experiments. It is composed of a motion table (XY), a pulse generator, an amplifier and a control PC. It injects EM pulses in a controlled way on a targeted device using an EM probe. It controls with a high precision the timing and the edges of the pulse. A recent development consists in adding a FPGA board to control the trigger in a more convenient and precise way. Then, the pulse can be triggered while a specific information is sent to the board under attack.

6.8.3. EMA

This last LHS platform is dedicated to side channel analysis (SCA) for evaluating the capabilities of dynamic countermeasure developed in the ANR Cogito. This platform uses a low cost oscilloscope, an EM probe and a set of software developed to perform the analysis. An additional oscilloscope, more performant has been added to the platform to target faster devices. We use the Julia language platform and custom developments to control the target and to analyze the EM traces.

7. New Results

7.1. Results for Axis 1: Vulnerability analysis

Statistical model checking employs Monte Carlo methods to avoid the state explosion problem of probabilistic (numerical) model checking. To estimate probabilities or rewards, SMC typically uses a number of statistically independent stochastic simulation traces of a discrete event model. Being independent, the traces may be generated on different machines, so SMC can efficiently exploit parallel computation. Reachable states are generated on the fly and SMC tends to scale polynomially with respect to system description. Properties may be specified in bounded versions of the same temporal logics used in probabilistic model checking. Since SMC is applied to finite traces, it is also possible to use logics and functions that would be intractable or undecidable for numerical techniques.

Several model checking tools have added SMC as a complement to exhaustive model checking. This includes the model checker UPPAAL, for timed automata, the probabilistic model checker PRISM, and the model checker Ymer, for continuous time Markov chains. Plasma Lab [29] is the first platform entirely dedicated to SMC. Contrary to other tools, that target a specific domain and offer several analysis techniques, including basic SMC algorithms, Plasma Lab is designed as a generic platform that facilitates multiple SMC algorithms, multiple modelling and query languages and has multiple modes of use. This allows us to apply statistical model checking techniques to a wide variety of problems, reusing existing simulators. With this process we avoid the task of rewriting a model of a system in a language not ideally design to do it. This complex task often leads either to an approximation of the original system or to a more complex model harder to analyze. The task needed to support a new simulator is to implement an interface plugin between our platform Plasma Lab and the existing tool, using the public API of our platform. This task has to be performed only once to analyze all the systems supported by the existing simulator.

Plasma Lab can already be used with the PRISM language for continuous and discrete time Markov chains and biological models. During the last years we have developed several new plugins to support SystemC language [50], Simulink models [70], dynamic software architectures language [41], [14], and train interlocking systems [64]. They have been presented in recent publications.

[50] Transaction-level modeling with SystemC has been very successful in describing the behavior of embedded systems by providing high-level executable models, in which many of them have an inherent probabilistic behavior, i.e., random data, unreliable components. It is crucial to evaluate the quantitative and qualitative analysis of the probability of the system properties. Such analysis can be conducted by constructing a formal model of the system and using probabilistic model checking. However, this method is unfeasible for large and complex systems due to the state space explosion. In this paper, we demonstrate the successful use of statistical model checking to carry out such analysis directly from large SystemC models and allows designers to express a wide range of useful properties.

[70] We present an extension of the statistical model checker Plasma Lab capable of analyzing Simulink models.

[41], Dynamic software architectures emerge when addressing important features of contemporary systems, which often operate in dynamic environments subjected to change. Such systems are designed to be reconfigured over time while maintaining important properties, e.g., availability,

correctness, etc. Verifying that reconfiguration operations make the system to meet the desired properties remains a major challenge. First, the verification process itself becomes often difficult when using exhaustive formal methods (such as model checking) due to the potentially infinite state space. Second, it is necessary to express the properties to be verified using some notation able to cope with the dynamic nature of these systems. Aiming at tackling these issues, we introduce DynBLTL, a new logic tailored to express both structural and behavioral properties in dynamic software architectures. Furthermore, we propose using statistical model checking (SMC) to support an efficient analysis of these properties by evaluating the probability of meeting them through a number of simulations. In this paper, we describe the main features of DynBLTL and how it was implemented as a plug-in for PLASMA, a statistical model checker.

[14] The critical nature of many complex software-intensive systems calls for formal, rigorous architecture descriptions as means of supporting automated verification and enforcement of architectural properties and constraints. Model checking has been one of the most used techniques to automatically analyze software architectures with respect to the satisfaction of architectural properties. However, such a technique leads to an exhaustive exploration of all possible states of the system under verification, a problem that becomes more severe when verifying dynamic software systems due to their typical non-deterministic runtime behavior and unpredictable operation conditions. To tackle these issues, we propose using statistical model checking (SMC) to support the analysis of dynamic software architectures while aiming at reducing effort, computational resources, and time required for this task. In this paper, we introduce a novel notation to formally express architectural properties as well as an SMC-based toolchain for verifying dynamic software architectures described in π-ADL, a formal architecture description language. We use a flood monitoring system to show how to express relevant properties to be verified, as well as we report the results of some computational experiments performed to assess the efficiency of our approach.

[64] , accepted at HASE 2017 In the railway domain, an interlocking is the system ensuring safe train traffic inside a station by controlling its active elements such as the signals or points. Modern interlockings are configured using particular data, called application data, reflecting the track layout and defining the actions that the interlocking can take. The safety of the train traffic relies thereby on application data correctness, errors inside them can cause safety issues such as derailments or collisions. Given the high level of safety required by such a system, its verification is a critical concern. In addition to the safety, an interlocking must also ensure that availability properties, stating that no train would be stopped forever in a station, are satisfied. Most of the research dealing with this verification relies on model checking. However, due to the state space explosion problem, this approach does not scale for large stations. More recently, a discrete event simulation approach limiting the verification to a set of likely scenarios, was proposed. The simulation enables the verification of larger stations, but with no proof that all the interesting scenarios are covered by the simulation. In this paper, we apply an intermediate statistical model checking approach, offering both the advantages of model checking and simulation. Even if exhaustiveness is not obtained, statistical model checking evaluates with a parameterizable confidence the reliability and the availability of the entire system.

7.1.1. Verification of Dynamic Software Architectures

Participants: Axel Legay, Jean Quilbeuf, Louis-Marie Traonouez.

Dynamic software architectures emerge when addressing important features of contemporary systems, which often operate in dynamic environments subjected to change. Such systems are designed to be reconfigured over time while maintaining important properties, e.g., availability, correctness, etc. π-ADL is a formal, well-founded theoretically language intended to describe software architectures under both structural and behavioral viewpoints. In order to cope with dynamicity concerns, π-ADL is endowed with architectural level primitives for specifying programmed reconfiguration operations, i.e., foreseen, pre-planned changes described at design time and triggered at runtime by the system itself under a given condition or event. Additionally, code source in the Go programming language is automatically generated from π-ADL architecture descriptions, thereby allowing for their execution.

We have developed with Plasma Lab a toolchain [14] for verifying dynamic software architectures described in π-ADL. The architecture description in π-ADL is translated towards generating source code in Go. As π-ADL architectural models do not have a stochastic execution, they are linked to a stochastic scheduler parameterized by a probability distribution for drawing the next action. Furthermore, we use existing probability distribution Go libraries to model inputs of system models as user functions. The program resulting from the compilation of the generated Go source code emits messages referring to transitions from addition, attachment, detachment, and value exchanges of architectural elements. Additionally we have introduced DynBLTL [41] a new logic tailored to express both structural and behavioral properties in dynamic software architectures.

We have developed two plugins atop the PLASMA platform, namely (i) a simulator plug-in that interprets execution traces produced by the generated Go program and (ii) a checker plugin that implements DynBLTL. With this toolchain, a software architect is able to evaluate the probability of a π-ADL architectural model to satisfy a given property specified in DynBLTL.

7.1.2. Statistical Model-Checking of Scheduling Systems

Participants: Axel Legay, Louis-Marie Traonouez.

Cyber-Physical Systems (CPS) are software implemented control systems that control physical objects in the real world. These systems are being increasingly used in many critical systems, such as avionics and automotive systems. They are now integrated into high performance platforms, with shared resources. This motivates the development of efficient design and verification methodologies to assess the correctness of CPS.

Schedulability analysis is a major problem in the design of CPS. Software computations that implements the commands sent to the CPS are split into a set of hard real-time tasks, often periodic. These tasks are associated to strict deadlines that must be satisfied. A scheduler is responsible for dispatching a shared resource (usually CPU computation time) among the different tasks according to a chosen scheduling policy. The schedulability analysis consists in verifying that the tasks always meet their deadlines.

Over the years, the schedulability of CPS have mainly been performed by analytical methods. Those techniques are known to be effective but limited to a few classes of scheduling policies. In a series of recent work, it has been shown that schedulability analysis of CPS could be performed with a model-based approach and extensions of verification tools such as UPPAAL. It shows that such models are flexible enough to embed various types of scheduling policies that go beyond those in the scope of analytical tools.

We have extended these works to include more complex features in the design of these systems and we have experimented the use of statistical model checking as a lightweight verification technique for these systems.

We also extended the approach to statistical model checking of products lines. Our first contribution has been to propose models to design software product lines (SPL) of preemptive real-time systems [25]. Software Product Line Engineering (SPLE) allows reusing software assets by managing the commonality and variability of products. Recently, SPLE has gained a lot of attention as an approach for developing a wide range of software products from non-critical to critical software products, and from application software to platform software products.

Real-time software products (such as real-time operating systems) are a class of systems for which SPLE techniques have not drawn much attention from researchers, despite the need to efficiently reuse and customize real-time artifacts. We have proposed a formal SPLE framework for real-time systems. It focuses on the formal analysis of real-time properties of an SPL in terms of resource sharing with time dependent functionalities. Our framework provides a structural description of the variability and the properties of a real time system, and behavioral models to verify the properties using formal techniques implemented in the tools UPPAAL symbolic model checker and UPPAAL statistical model checker. For the specification of an SPL, we propose an extension of a feature model, called Property Feature Model (PFM). A PFM explicitly distinguishes features and properties associated with features, so that properties are analyzed in the context of the relevant features. We also define a non-deterministic decision process that automatically configures the products of an SPL that satisfy the constraints of a given PFM and the product conditions of customers. Finally we analyze the products against the associated properties. For analyzing real-time properties, we provide feature behavioral models of the components of a scheduling unit, i.e. tasks, resources and schedulers. Using these feature behavioral models, a family of scheduling units of an SPL is formally analyzed against the designated properties with model checking techniques.

[25] This paper presents a formal analysis framework to analyze a family of platform products w.r.t. real-time properties. First, we propose an extension of the widely-used feature model, called Property Feature Model (PFM), that distinguishes features and properties explicitly. Second, we present formal behavioral models of components of a real-time scheduling unit such that all real-time scheduling units implied by a PFM are automatically composed to be analyzed against the properties given by the PFM. We apply our approach to the verification of the schedulability of a family of scheduling units using the symbolic and statistical model checkers of UPPAAL.

7.1.3. Model-based Framework for Hierarchical Scheduling Systems

Participants: Axel Legay, Louis-Marie Traonouez, Mounir Chadli.

In order to reduce costs in the design of modern CPS, manufacturers devote strong efforts to maximize the number of components that can be integrated on a given platform. This can be achieved by minimizing the resource requirements of individual components. A hierarchical scheduling systems (HSS) integrates a number of components into a single system running on one execution platform. Hierarchical scheduling systems have been gaining more attention by automotive and aircraft manufacturers because they are practical in minimizing the cost and energy of operating applications.

Several papers have proposed model-based compositional framework for HSS. In [4] we proposed a methodology for optimizing the resource requirement of a component of an HSS using model checking techniques. Our methodology consists of using a lightweight statistical model checking method and a costly but absolute certain symbolic model checking method that operates on identical models.

In another work [15] we have proposed stochastic extension of HSS that allows us to capture tasks whose real-time attributes, such as deadline, execution time or period, are also characterized by probability distributions. This is particularly useful to describe mixed-critical systems and make assumptions on the hardware domain. These systems combine hard real-time periodic tasks, with soft real-time sporadic tasks. Classical scheduling techniques can only reason about worst case analysis of these systems, and therefore always return pessimistic results. Using tasks with stochastic period we can better quantify the occurrence of these tasks. Similarly, using stochastic deadlines we can relax timing requirements, and stochastic execution times are used to model the variation of the computation time needed by the tasks. These distributions can be sampled from executions or simulations of the system, or set as requirements from the specifications. For instance in avionics, display components have lower criticality. They can include sporadic tasks generated by users requests. Average user demand is efficiently modeled with a probability distribution.

We have also developed a graphical high-level language to represent scheduling units and complex hierarchical scheduling systems. In order to bridge the gap between the formalisms, we exploit Cinco, a generator for domain specific modeling tools to generate an interface between this language and the one of UPPAAL. Cinco allows to specify the features of a graphical interface in a compact meta-model language. This is a flexible approach that could be extended to any formal model of scheduling problem.

[4] Compositional reasoning on hierarchical scheduling systems is a well-founded formal method that can construct schedulable and optimal system configurations in a compositional way. However, a compositional framework formulates the resource requirement of a component, called an interface, by assuming that a resource is always supplied by the parent components in the most pessimistic way. For this reason, the component interface demands more resources than the amount of resources that are really sufficient to satisfy sub-components. We provide two new supply bound functions which provides tighter bounds on the resource requirements of individual components. The tighter bounds are calculated by using more information about the scheduling system. We evaluate our new tighter bounds by using a model-based schedulability framework for hierarchical scheduling systems realized as UPPAAL models. The timed models are checked using model checking tools UPPAAL and UPPAAL SMC, and we compare our results with the state of the art tool CARTS.

[15] Over the years, schedulability of Cyber-Physical Systems (CPS) have mainly been performed by analytical methods. Those techniques are known to be effective but limited to a few classes of scheduling policies. In a series of recent work, we have shown that schedulability analysis of CPS could be performed with a model-based approach and extensions of verification tools such as UPPAAL. One of our main contribution has been to show that such models are flexible enough to embed various types of scheduling policies that go beyond those in the scope of analytical tools. In this paper, we go one step further and show how our formalism can be extended to account for stochastic information, such as sporadic tasks whose attributes depend on the hardware domain. Our second contribution is to make our tools accessible to average users that are not experts in formal methods. For doing so, we propose a graphical and user-friendly language that allows us to describe scheduling problems. This language is automatically translated to formal models by exploiting a meta-model approach. The principle is illustrated on a case study.

7.1.4. Verification of Interlocking Systems

Participants: Axel Legay, Louis-Marie Traonouez, Jean Quilbeuf.

An interlocking is a system that controls the train traffic by acting as an interface between the trains and the railway track components. The track components are for example, the signals that allow the train to proceed, or the points that guide the trains from one track to another. The paths followed by the trains are called routes. Modern interlockings are computerized systems that are composed of generic software and application data.

We have proposed in collaboration with Université Catholique de Louvain and Alstom a method to automatically verify an interlocking using simulation and statistical model checking [64]. We use a simulator developed by Université Catholique de Louvain that is able to generate traces of the interlocking systems from a track layout and application data. This simulator is plug with Plasma Lab using a small interface developed with Plasma Lab's API. Then, the traces generated by the simulator have been used by Plasma Lab SMC algorithms to measure the correctness of the system. We have used Monte-Carlo and importance splitting algorithms to verify this system.

7.1.5. Advanced Statistical Model Checking

Participants: Axel Legay, Sean Sedwards, Louis-Marie Traonouez.

Statistical model checking (SMC) addresses the state explosion problem of numerical model checking by estimating quantitative properties using simulation. Rare events, such as software bugs, are often critical to the performance of systems but are infrequently observed in simulations. They are therefore difficult to quantify using SMC. Nondeterministic systems deliberately leave parts of system behaviour undefined, hence it is not immediately possible to simulate them. Our ongoing work thus pushes the boundaries of the cutting edge of SMC technology by focusing on rare event verification and the optimisation of nondeterminism.

7.1.5.1. Optimizing Nondeterministic Systems

[17] Probabilistic timed automata (PTA) generalize Markov decision processes (MDPs) and timed automata (TA), both of which include nondeterminism. MDPs have discrete nondeterministic choices, while TA have continuous nondeterministic time. In this work we consider finding *schedulers* that resolve all nondeterministic choices in order to maximize or minimize the probability of a time-bounded LTL property. Exhaustive numerical approaches often fail due to state explosion, hence we present a new lightweight on-the-fly algorithm to find near-optimal schedulers. To discretize the continuous choices we make use of the classical region and zone abstractions from timed automata model checking. We then apply our recently developed "smart sampling" technique for statistical verification of Markov decision processes. On standard case studies our algorithm provides good estimates for both maximum and minimum probabilities. We compare our new approach with alternative techniques, first using tractable examples from the literature, then motivate its scalability using case studies that are intractable to numerical model checking and challenging for existing statistical techniques.

7.1.5.2. Rare Event Verification

[3] Importance sampling is a standard technique to significantly reduce the computational cost of quantifying rare properties of probabilistic systems. It works by weighting the original distribution of the system to make the rare property appear more frequently in simulations, then compensating the resulting estimate by the weights. This can be done on the fly with minimal storage, but the challenge is to find *near optimal* importance sampling distributions efficiently, where optimal means that paths that do not satisfy the property are never seen, while paths that satisfy the property appear in the same proportion as in the original distribution.

Our approach uses a tractable cross-entropy minimization algorithm to find an optimal parameterized importance sampling distribution. In contrast to previous work, our algorithm uses a naturally defined low dimensional vector to specify the distribution, thus avoiding an explicit representation of a transition matrix. Our parametrisation leads to a unique optimum and is shown to produce many orders of magnitude improvement in efficiency on various models. In this work we specifically link the existence of optimal importance sampling distributions to time-bounded logical properties and show how our parametrisation affects this link. We also motivate and present simple algorithms to create the initial distribution necessary for cross-entropy minimization. Finally, we discuss the open challenge of defining error bounds with importance sampling and describe how our optimal parameterized distributions may be used to infer qualitative confidence.

[10] In this work we consider rare events in systems of Stochastic Timed Automata (STA) with time-bounded reachability properties. This model may include rarity arising from explicit discrete transitions, as well as more challenging rarity that results from the intersection of timing constraints and continuous distributions of time. Rare events have been considered with simple combinations of continuous distributions before, e.g., in the context of queueing networks, but we present an automated framework able to work with arbitrarily composed STA. By means of symbolic exploration we first construct a zone graph that excludes unfeasible times. We then simulate the system within the zone graph, avoiding "dead ends" on the fly and proportionally redistributing their probability to feasible transitions. In contrast to many other importance sampling approaches, our "proportional dead end avoidance" technique is guaranteed by construction to reduce the variance of the estimator with respect to simulating the original system. Our results demonstrate that in practice we can achieve substantial overall computational gains, despite the symbolic analysis.

[49] In this invited paper we outline some of our achievements in quantifying rare properties in the context of SMC. In addition to the importance sampling techniques described above, we also describe our work on importance *splitting*. Importance splitting works by decomposing the probability of a rare property into a product of probabilities of sub-properties that are easier to estimate. The sub-properties are defined by *levels* of a *score function* that maps states of the system$\times$property product automaton to values. We have provided the first general purpose implementation of this approach, using user-accessible "observers" that are compiled automatically from the property. These observers may be used by both fixed and adaptive level importance splitting algorithms and are specifically designed to make distribution efficient.

7.1.6. Side-channel Analysis of Cryptographic Substitution Boxes

Participants: Axel Legay, Annelie Heuser.

With the advent of the Internet of Things, we are surrounded with smart objects (aka things) that have the ability to communicate with each other and with centralized resources. The two most common and widely noticed artefacts are RFID and Wireless Sensor Networks which are used in supply-chain management, logistics, home automation, surveillance, traffic control, medical monitoring, and many more. Most of these applications have the need for cryptographic secure components which inspired research on cryptographic algorithms for constrained devices. Accordingly, lightweight cryptography has been an active research area over the last 10 years. A number of innovative ciphers have been proposed in order to optimize various performance criteria and have been subject to many comparisons. Lately, the resistance against side-channel attacks has been considered as an additional decision factor.

Side-channel attacks analyze physical leakage that is unintentionally emitted during cryptographic operations in a device (e.g., power consumption, electromagnetic emanation). This side-channel leakage is statistically dependent on intermediate processed values involving the secret key, which makes it possible to retrieve the secret from the measured data.

Side-channel analysis (SCA) for lightweight ciphers is of particular interest not only because of the apparent lack of research so far, but also because of the interesting properties of substitution boxes (S-boxes). Since the nonlinearity property for S-boxes usually used in lightweight ciphers (i.e., 4×4) can be maximally equal to 4, the difference between the input and the output of an S-box is much smaller than for instance for AES. Therefore, one could conclude that from that aspect, SCA for lightweight ciphers must be more difficult. However, the number of possible classes (e.g., Hamming weight (HW) or key classes) is significantly lower, which may indicate that SCA must be easier than for standard ciphers. Besides the difference in the number of classes and consequently probabilities of correct classification, there is also a huge time and space complexity advantage (for the attacker) when dealing with lightweight ciphers.

In [23], [67] we give a detailed study of lightweight ciphers in terms of side-channel resistance, in particular for software implementations. As a point of exploitation we concentrate on the non-linear operation (S-box) during the first round. Our comparison includes SPN ciphers with 4-bit S-boxes such as KLEIN, PRESENT, PRIDE, RECTANGLE, Mysterion as well as ciphers with 8-bit S-boxes: AES, Zorro, Robin. Furthermore, using simulated data for various signal-to-noise ratios (SNR) we present empirical results for Correlation Power Analysis (CPA) and discuss the difference between attacking 4-bit and 8-bit S-boxes.

Following this direction current studies evaluate and connect cryptographic properties with side-channel resistance. More precisely, in an ideal setting a cipher should be resilient against cryptanalyses as well as side-channel attacks and yet easy and cheap to be implemented. However, since that does not seem to be possible at the current level of knowledge, we are required to make a number of trade-offs. Therefore, we investigate several S-box parameters and connect them with well known cryptographic properties of S-boxes. Moreover, when possible we give clear theoretical bounds on those parameters as well as expressions connecting them with properties like nonlinearity and δ-uniformity. We emphasize that we select the parameters to explore on the basis of their possible connections with the side-channel resilience of S-boxes.

To this end, we divide the primary goal into several sub-problems. First, we discuss what is the maximal number of fixed points one can have in an optimal S-box. The question of the maximal number of fixed points for an optimal S-box is of interest on its own, but also in the side-channel context since intuitively an S-box with many fixed points should consume less power and therefore have less leakage. Moreover, the preservation of Hamming weight and a small Hamming distance between x and $F(x)$ are two more properties each of which could strengthen the resistance to SCA. Our findings show that notably in the case when exactly preserving the Hamming weight, the confusion coefficient reaches good value and consequently the S-box has good SCA resilience. We show that the S-boxes with no differences in the Hamming weight of their input and output pairs (and even, S-boxes F such that $F(x)$ have Hamming weight near the Hamming weight of x, on average) or S-boxes such that $F(x)$ lies at a small Hamming distance from x cannot have high nonlinearity (although the obtainable values are not too bad for $n = 4, 8$) and therefore are not attractive in practical applications. Note that an S-box with many fixed points is also a particular case of an S-box that preserves the Hamming weight/distance between the inputs and outputs. Furthermore, our study includes involutive functions since they have a particular advantage over general pseudo-permutations. In particular, not only from an implementation viewpoint but also their side-channel resilience is the same regardless if an attacker considers the encryption or decryption phase as well as attacking the first or the last round. Next, we find a theoretical expression connecting the confusion coefficient with that of preserving the Hamming weight of inputs and outputs.

In the practical part, we first confirm our theoretical findings about the connection between preserving Hamming weight and the confusion coefficient. Besides that, we give a number of S-box examples of size 4×4 intended to provide more insight into possible trade-offs between cryptographic properties and side-channel resilience. However, our study shows that mostly preserving Hamming weight might not automatically result in a small minimum confusion coefficient and thus in higher side-channel resistance. We therefore in detail examine the influence of F on the confusion coefficient in general by concentrating on the input (in which key hypothesis are made) and the minimum value of the confusion coefficient. Following, we evaluate a number of S-boxes used in today's ciphers and show that their SCA resilience can significantly differ. Finally, we point out that non-involutive S-boxes might bring a significant disadvantage in case an attacker combines the information about F and F^{-1} by either targeting both first and last round of an algorithm or encryption and decryption.

[67] Side-channel Analysis of Lightweight Ciphers: Current Status and Future Directions

[23] Side-channel Analysis of Lightweight Ciphers: Does Lightweight Equal Easy?

7.1.7. Binary Code Analysis: Formal Methods for Fault Injection Vulnerability Detection

Participants: Axel Legay, Thomas Given-Wilson, Nisrine Jafri, Jean-Louis Lanet.

Formal methods such as model checking provide a powerful tool for checking the behaviour of a system. By checking the properties that define correct system behaviour, a system can be determined to be correct (or not).

Increasingly fault injection is being used as both a method to attack a system by a malicious attacker, and to evaluate the dependability of the system. By finding fault injection vulnerabilities in a system, the resistance to attacks or faults can be detected and subsequently addressed.

A process is presented that allows for the automated simulation of fault injections. This process proceeds by taking the executable binary for the system to be tested, and validating the properties that represent correct system behaviour using model checking. A fault is then injected into the executable binary to produce a mutant binary, and the mutant binary is model checked also. A different result to the validation of the executable binary in the checking of the mutant binary indicates a fault injection vulnerability.

This process has been automated with existing tools, allowing for easy checking of many different fault injection attacks and detection of fault injection vulnerabilities. This allows for the detection of fault injection vulnerabilities to be fully automated, and broad coverage of the system to be formally shown.

7.1.8. Security at the hardware and software boundaries

Participants: Axel Legay, Nisrine Jafri, Jean-Louis Lanet, Ronan Lashermes, Hélène Le Bouder.

7.1.8.1. IoT security

IoT security has to face all the challenges of the mainstream computer security but also new threats. When an IoT device is deployed, most of the time it operates in a hostile environment, i.e. the attacker can perform any attack on the device. If secure devices use tamper resistant chip and are programmed in a secure manner, IoT use low cost micro-controllers and are not programmed in a secure way. We developed new attacks but also evaluate how the code polymorphism can be used against these attacks. In [45] [27] we developed a template attack to retrieve the value of a PIN code from a cellphone. We demonstrated that the maximum trials to retrieve the four bytes of secret PIN is 8 and in average 3 attempts are sufficient. A supervised learning algorithm is used.

Often smart phones allow up to 10 attempts before locking definitely the memory. We used an embedded code generator [16], [45] dedicated to a given security function using a DSL to increase the security level of a non tamper resistant chip. We brought to the fore that a design of the software for protecting against fault attacks decreases the security against SCA. Fault attack is a mean to execute a code that is slightly different from the one that has been loaded into the device. Thus, to be sure that a genuine code cannot be dynamically transformed, one needs to analyze any possibility of a code to be transformed.

The work presented in [34] made possible to design an extremely effective architecture to achieve Montgomery modular multiplication. The proposed solution combines a limited resource consumption with the lowest latency compared with the literature. This allows to envisage new applications of asymmetric cryptography in systems with few resources. In order to find a cryptographic key using hidden channels, most attacks use the a priori knowledge of texts sent or received by the target. The proposed analysis presented in [28] does not use these assumptions. A belief propagation technique is used to cross the information obtained from leaked information with the equations governing the targeted algorithm.

7.1.8.2. Safe update mechanism for IoT

One of the challenges for IoT is the possibility to update the code through a network. This is done by stopping the system, loading the new version, verifying the signature of the firmware and installing it into the memory. Then, the memory must be cleaned to eliminate the code and the data of the previous version. Some IoT (sensor acquisition and physical system control) requires to never stop while executing the code. We have developed a complete architecture that performs such an update with real time capabilities. If one wants to use this characteristic in a real world it should pass certification. In particular he has to demonstrate that the system performs as expected. We used formal methods (mainly Coq) to demonstrate that the semantics of the code is preserved during the update. In [30], we paid attention to the detection of the Safe Update Point (SUP) because our implementation had some time an unstable behavior. We demonstrated that in a specific case, while several threads using code to be updated, the system enters into a deadlock. After discovering the bug, we patched our system.

7.1.8.3. Reverse engineering of firmware

Reverse engineering has two aspects; code reverse for which the literature is abundant and data reverse i.e. understanding the meaning of a structure and its usage has been less studied. The first step in reverse engineering consists in getting access of the code. In the case of romized code in a SoC, the access to the code is protected by a MMU mechanism and thus is an efficient mitigation mechanism against reverse engineering. In [8], [2] and [33] we set up several attacks to get access to the code even in presence of a MMU. The attack in [8] uses a vulnerability in the API where an object can be used instead of an array. This allows to read and write the code area leading to the possibility to execute arbitrary code in memory. In [33], we use the attack tree paradigm to explore all the possibilities to mount an attack on a given product. In [2], we used a ROP (Return Oriented Programming) attack to inject a shell code in the context of another application. Due to the fact that the shell code is executed in the context of the caller, the firewall is unable to detect the access to the secure container of the targeted application. This allows us to retrieve the content of the secure containers.

Once the dump has been obtained, one can try to retrieve code and data. Retrieving code is not obvious but several tools exist to help the analyst. These tools require that all the ISA (Instruction Set Architecture) is known. Sometime, the ISA is not known and in particular when one wants to obfuscate the code, he can use a virtual machine to execute dedicated byte code. In [32], we developed a methodology to infer the missing byte code, then we execute a data type inference to understand the memory management algorithm.

7.2. Results for Axis 2: Malware analysis

The detection of malicious programs is a fundamental step to be able to guarantee system security. Programs that exhibit malicious behavior, or *malware*, are commonly used in all sort of cyberattacks. They can be used to gain remote access on a system, spy on its users, exfiltrate and modify data, execute denial of services attacks, etc.

Significant efforts are being undertaken by software and data companies and researchers to protect systems, locate infections, and reverse damage inflicted by malware. Malware analysis can be divided in the following three main problems:

7.2.1. Malware Detection

Participants: Axel Legay, Fabrizio Biondi, Olivier Decourbe, Mike Enescu, Thomas Given-Wilson, Annelie Heuser, Nisrine Jafri, Jean-Louis Lanet, Jean Quilbeuf.

Given a file or data stream, the malware detection problem consists of understanding if the file or data stream contain traces of malicious behavior. For binary executable files in particular, this requires reverse engineering the file's behavior to understand if it is malicious. The main reverse engineering techniques are categorized as:

Static Analysis This refers to techniques that analyze the file without executing it. It includes disassembling the file's executable code and analyzing other static features of the binary, like its import/export table, hash, etc. The file's control flow and system flow graphs can be retrieved statically (unless they are obfuscated; see below) and used to guide the exploration of the file's semantics in the search of malicious behavior. Information flow can be tracked since hostile applications often try to transmit private information to distant servers (this form of malware are now widely spread in the mobile world). The challenge consists in detecting into a file that a private information does not leak to the external world. The verification can be done statically, dealing with storage channel (implicit or explicit), but not with side channel.

Dynamic Analysis This refers to techniques that actually executed the file in a sandbox (usually a virtualized environment) and analyze its interaction with the sandbox. This technique is effective in understanding the file's actual interactions with the system, making it easy to detect malicious behavior. However, malware often implements sandbox detection techniques to detect when it is being run in a virtualized environment, when functions or system calls are hooked by the analyst, or when the sandbox does not look like a normal user's machine (e.g. because it does not contain any document). Dynamic tracking of information flow makes it possible to cope with side channel attacks. With temporal side channel, the challenge lies in the potential declassification procedure used by malware to escape the analysis. We extend the TaintDroid framework to cope with native code invocation [47]. This approach reduces the false positive warning drastically. Recently we have extended this work to cope with timing side channels [under submission]. We are developing a new malware that declassifies the labels thanks to the audio system of the smart-phone. This is a joint work with Telecom Bretagne.

Hybrid Analysis This refers to technique that combine both static and dynamic behavior, i.e. both code analysis and execution. While more complex to implement, these techniques are able to overcome many of the shortcomings of full static and full dynamic analysis. The best example of a hybrid technique is concolic (a portmanteau for CONCrete + symbOLIC) analysis.

To contribute to concolic analysis, we are working on the state-of-the-art `angr` concolic execution engine to make it fast and efficient enough to analyze large executable malware files efficiently. We are improving `angr` 's parallelism and allowing it to precompute semantic stubs of function and system calls, allowing it to focus its analysis on the main file without having to branch in the rest of the operative system. We plan to contribute our improvements to the main `angr` branch, so that the whole community can benefit from them.

7.2.2. *Malware Deobfuscation*

Participants: Axel Legay, Fabrizio Biondi, Olivier Decourbe, Mike Enescu, Thomas Given-Wilson, Annelie Heuser, Nisrine Jafri, Jean-Louis Lanet, Jean Quilbeuf.

Given a file (usually a portable executable binary or a document supporting script macros), deobfuscation refers to the preparation of the file for the purposes of further analysis. Obfuscation techniques are specifically developed by malware creators to hinder detection reverse engineering of malicious behavior. Some of these techniques include:

Packing Packing refers to the transformation of the malware code in a compressed version to be dynamically decompressed into memory and executed from there at runtime. Packing techniques are particularly effective against static analysis, since it is very difficult to determine statically the content of the unpacked memory to be executed, particularly if packing is used multiple times. The compressed code can also be encrypted, with the key being generated in a different part of the code and used by the unpacking procedure, or even transmitted remotely from a command and control (C&C) server.

Control Flow Flattening This technique aims to hinder the reconstruction of the control flow of the malware. The malware's operation are divided into basic blocks, and a dispatcher function is created that calls the blocks in the correct order to execute the malicious behavior. Each block after its execution returns control to the dispatcher, so the control flow is flattened to two levels: the dispatcher above and all the basic blocks below.

To prevent reverse engineering of the dispatcher, it is often implemented with a cryptographic hash function. A more advanced variant of this techniques embed a full virtual machine with a randomly generated instruction set, a virtual program counted, and a virtual stack in the code, and uses the machine's interpreter as the dispatcher.

Virtualization is a very effective technique to prevent reverse engineering. To contrast it, we are implementing state-of-the-art devirtualization algorithms in `angr` , allowing it to detect and ignore the virtual machine code and retrieving the obfuscated program logic. Again, we plan to contribute our improvements to the main `angr` branch, thus helping the whole security community fighting virtualized malware.

Opaque Constants and Conditionals Reversing packing and control flow flattening techniques requires understanding of the constants and conditionals in the program, hence many techniques are deployed to obfuscate them and make them unreadable by reverse engineering techniques. Such techniques are used e.g. to obfuscate the decryption keys of packed encrypted code and the conditionals in the control flow.

We have proven the efficiency of dynamic synthesis in retrieving opaque constant and conditionals, compared to the state-of-the-art approach of using SMT (Satisfiability Modulo Theories) solvers, when the input space of the opaque function is small enough. We are developing techniques based on fragmenting and analyzing by brute force the input space of opaque conditionals, and SMT constraints in general, to be integrated in SMT solvers to improve their effectiveness.

7.2.3. Malware Classification

Participants: Axel Legay, Fabrizio Biondi, Olivier Decourbe, Mike Enescu, Thomas Given-Wilson, Annelie Heuser, Nisrine Jafri, Jean-Louis Lanet, Jean Quilbeuf.

Once malicious behavior has been located, it is essential to be able to classify the malware in its specific family to know how to disinfect the system and reverse the damage inflicted on it.

While it is rare to find an actually previously unknown malware, morphic techniques are employed by malware creators to ensure that different generations of the same malware behave differently enough than it is hard to recognize them as belonging to the same family. In particular, techniques based on the syntax of the program fails against morphic malware, since syntax can be easily changed.

To this end, semantic signatures are used to classify malware in the appropriate family. Semantic signatures capture the malware's behavior, and are thus resistant to morphic and differentiation techniques that modify the malware's syntactic signatures. We are investigating semantic signatures based on the program's System Call Dependency Graph (SCDG), which have been proven to be effective and compact enough to be used in practice. SCDGs are often extracted using a technique based on pushdown automata that is ineffective against obfuscated code; instead, we are applying concolic analysis via the `angr` engine to improve speed and coverage of the extraction.

Once a semantic signature has been extracted, it has to be compared against large database of known signatures representing the various malware families to classify it. The most efficient way to obtain this is to use a supervised machine learning classifier. In this approach, the classifier is trained with a large sample of signatures malware annotated with the appropriate information about the malware families, so that it can learn to quickly and automatically classify signatures in the appropriate family. Our work on machine learning classification focuses on using SCDGs as signatures. Since SCDGs are graphs, we are investigating and adapting algorithms for the machine learning classification of graphs, usually based on measures of shared subgraphs between different graphs.

In malware detection and classification, it is fundamental to have a false positive rate (i.e. rate of cleanware classified as malware) approaching zero, otherwise the classification system will classify hundred or thousands of cleanware files as malware, making it useless in practice. To decrease the false positive rate, the classifier is also trained with a large and representative database of cleanware, so that it can discriminate between signatures of cleanware and malware with a minimal false positive rate. We use a large database of malware and cleanware to train our classifier, thus guaranteeing a high detection rate with a small false positive rate.

7.2.4. Papers

This section gathers papers that are results common to all sections above pertaining to Axis 2.

[57] Black-box synthesis is more efficient than SMT deobfuscation on predicates obfuscated with Mixed-Boolean Arithmetics.

[66] Recently fault injection has increasingly been used both to attack software applications, and to test system robustness. Detecting fault injection vulnerabilities has been approached with a variety of methods, yielding varied results. This paper proposes a general process using model checking to detect fault injection vulnerabilities in binaries. The process is implemented and used to detect a variety of different kinds of fault injection vulnerabilities in binaries.

[59] Fault-injection exploits hardware weaknesses to perturbate the behaviour of embedded devices. Here, we present new model-based techniques and tools to detect such attacks developed at the High-Security Laboratory at Inria.

[52] We proposed to use a bare metal approach without virtualization and a method to let the system stop the execution while the malware has been deployed in memory.

[51] We present our framework to grab sample from the net, evaluate it on victim PC and detect its presence thanks to our counter measures.

[53] In this paper, two counter measures are presented. The first one is related with the mode ECB of the AES cryptographic algorithm and the second is related with the usage of the crypto API. We developed a cryptographic provider which intercepts the key generation and store it in a safe place. Then we are able to decipher any files that the malware should have encrypted.

7.3. Results for Axis 3: Building a secure network stack

7.3.1. Private set intersection cardinality

Participants: Jeffrey Burdges, Alvaro Garcia Recuero, Christian Grothoff.

We designed new efficient protocol for privacy-preserving signed set intersection cardinality using blinded BLS signatures over bilinear maps and demonstrated its utility in machine learning for abuse detection in decentralised online social networks. The paper was presented at DPM 2016 [21].

7.3.2. Cell tower privacy

Participants: Christian Grothoff, Neal Walfield.

We analyzed real-world mobility data based on cell tower traces, and illustrated how cell tower trace data can be used to identify patterns of life. We then used these results to predict future locations over a 24h period in 15 minute intervals with 80% accuracy [43].

7.3.3. Taler protocol improvements

Participants: Jeffrey Burdges, Florian Dold, Christian Grothoff, Marcello Stanisci.

We improved the Taler payment system protocol [13] to (1) reduce storage requirements for the exchange, which was the dominant cost, and (2) reduce security assumptions by avoiding the use of AES entirely.

We adapted the payment handshake to work even if JavaScript is disabled for the Web page, and adjusted the protocol to match discussions for future Web payment protocols from W3c. The protocol was extended with accounting functions to allow merchants to trace payments for their back office requirements. The user interface of the Taler wallet was streamlined, the wallet can finally get change, and the extension was made to work with Firefox. A public demonstrator was launched at https://demo.taler.net/.

7.4. Other research results: Information-Theoretical Quantification of Security Properties

Participants: Axel Legay, Fabrizio Biondi, Mounir Chadli, Thomas Given-Wilson.

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

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

Armed with the produced output of the system, the attacker tries to infer information about the secret information that produced the output. The quantitative analysis we consider defines and computes how much information the attacker can expect to infer (typically measured in bits). This expected leakage of bits is the information leakage of the system.

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

Applied to shared-key cryptosystems, this approach allows precise reasoning about the information leakage of the secret key when the attacker knows the encoder function and information about the distribution of messages. In such scenarios, this work has generalised perfect secrecy, and so provides a more useful measure for unconditional cryptosystems (results that are safe against future advances in computing capabilities and theoretical breakthroughs in unsolved problems).

This work also explored scenarios where the attacker has less information about the cryptosystem; such as not knowing the encoder function, or not knowing the message distribution. Results here formalised that the attacker can never improve their attacks by having bad prior information, thus ensuring misinformation is always useful. Also, results show that the choice of encoder function may strengthen the cryptosystem against being learned by the attacker through observation. In particular, we showed that a well designed encoder function (represented as a matrix) has an infinitude of freedom for the attacker. Thus, the attacker cannot accurately learn all the secret information merely by observation.

There are several different scenarios where the attacker is trying to learn the secret information about the system. Here this is explored by considering what the secret information is, or equivalently, what prior knowledge the attacker has about the system.

Our new results in information leakage computation include implementing a hybrid precise-statistical computation algorithm for our QUAIL tool. The new algorithm bridges the gap between statistical and formal techniques by using static program analysis to extract structural information about the program to be analyze and decide whether each part of it would be analyzed more efficiently with precise or statistical analysis. Then each part is analyzed with the most appropriate technique, and all analyses are combined into a final result. This new hybrid method outperforms precise and statistical analysis in computation time and precision, and is a clear example of the advantages of combining precise and statistical techniques.We refer to the tools section for more details.

Additionally, we have considered how the scheduling of privileged and unprivileged processes on a shared memory could allow an unprivileged process to access confidential information temporarily stored in the memory by a privileged process. This is for instance the case in cache attacks. We have developed a general model of information leakage for scheduled systems. Our model considers a finer granularity than previous attempts on the subject, allowing us to schedule processes with small leakage, and schedule sets of processes that were considered unschedulable with no leakage by the state of the art.

[1] Preserving the privacy of private communication is a fundamental concern of computing addressed by encryption. Information-theoretic reasoning models unconditional security where the strength of the results does not depend on computational hardness or unproven results. Usually the information leaked about the message by the ciphertext is used to measure the privacy of a communication , with perfect secrecy when the leakage is 0. However this is hard to achieve in practice. An alternative measure is the equivocation, intuitively the average number of message/key pairs that could have produced a given cipher-text. We show a theoretical bound on equivocation called max-equivocation and show that this generalizes perfect secrecy when achievable, and provides an alternative measure when perfect secrecy is not achievable. We derive bounds for max-equivocation for symmetric encoder functions and show that max-equivocation is achievable when the entropy of the ciphertext is minimized. We show that max-equivocation easily accounts for key re-use scenarios, and that large keys relative to the message perform very poorly under equivocation. We study encoders under this new perspective, deriving results on their achievable maximal equivocation and showing that some popular approaches such as Latin squares are not optimal. We show how unicity attacks can be naturally modeled, and how relaxing encoder symmetry improves equivocation. We present some algorithms for generating encryption functions that are practical and achieve 90 to 95% of the theoretical best, improving with larger message spaces.

[24] Analysis of a probabilistic system often requires to learn the joint probability distribution of its random variables. The computation of the exact distribution is usually an exhaustive precise analysis on all executions of the system. To avoid the high computational cost of such an exhaustive search, statistical analysis has been studied to efficiently obtain approximate estimates by analyzing only a small but representative subset of the system's behavior. In this paper we propose a hybrid statistical estimation method that combines precise and statistical analyses to estimate mutual information and its confidence interval. We show how to combine the analyses on different components of the system with different precision to obtain an estimate for the whole system. The new method performs weighted statistical analysis with different sample sizes over different components and dynamically finds their optimal sample sizes. Moreover it can reduce sample sizes by using prior knowledge about systems and a new abstraction-then-sampling technique based on qualitative analysis. We show the new method outperforms the state of the art in quantifying information leakage.

[12] The protection of users' data conforming to best practice and legislation is one of the main challenges in computer science. Very often, large-scale data leaks remind us that the state of the art in data privacy and anonymity is severely lacking. The complexity of modern systems make it impossible for software architect to create secure software that correctly implements privacy policies without the help of automated tools. The academic community needs to invest more effort in the formal modeling of security and anonymity properties, providing a deeper understanding of the underlying concepts and challenges and allowing the creation of automated tools to help software architects and developers. This research track provides numerous contributions to the formal modeling of security and anonymity properties and the creation of tools to verify them on large-scale software projects.

[62] High-security processes typically have to load confidential information, such as encryption keys or private data, into memory as part of their operation. In systems with a single shared memory, when high-security processes are switched out due to context switching, confidential information may remain in memory and be accessible to low-security processes. This paper considers this problem from the perspective of scheduling. A formal model supporting preemption is introduced that allows: reasoning about leakage between high-and low-security processes, and producing information-leakage aware schedulers. Several information-leakage aware heuristics are presented in the form of compositional pre-and postprocessors as part of a more general scheduling approach. The effectiveness of such heuristics is evaluated experimentally, showing them to achieve significantly better schedulability than the state of the art.

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Contracts with Industry

- CISCO (http://www.cisco.com)
- Thales (https://www.thalesgroup.com)
- Oberthur Technologies (http://www.oberthur.com/)

9. Partnerships and Cooperations

9.1. Regional Initiatives

ARED grant for Nisrine Jafri.

Postdocs grants for Fabrizio Biondi, Jeffrey Paul Burdges, Florian Dold, Ronan Lashermes.

9.2. National Initiatives

9.2.1. ANR

- ANR MALTHY, Méthodes ALgèbriques pour la vérification de modèles Temporisés et HYbrides, Thao Dang, 4 years, Inria and VISEO and CEA and VERIMAG
- ANR COGITO, Runtime Code Generation to Secure Devices,, 3 years, Inria and CEA and ENSMSE and XLIM.

9.3. European Initiatives

9.3.1. FP7 & H2020 Projects

9.3.1.1. ACANTO

Participants: Axel Legay, Thomas Given-Wilson, Sean Sedwards, Olivier Zendra.

Start: 2015. End: 2018.

The population of the advanced countries is ageing. This simple and widely recognised fact has important implications for health, society and economics. The most evident is in the number of people who report activity limitations, which grows significantly with age as clearly shown in the following chart. Activity limitations have an adverse effect on a person's productivity, on the quality of her social relations and, ultimately, on her quality of life. Policy makers confronted with a problem of challenging complexity: how to develop an effective strategy to fight the physical and cognitive decline of older adults in the face of ever shrinking financial resources for health care and social services.

In this context, technology can be of considerable help to care–givers to extend the range and the efficacy of their actions. The ACANTO project (http://www.ict-acanto.eu) aims to develop a portfolio of technical solution that can serve this purpose. More specifically, our goal is to spur older adults into a sustainable and regular level of physical exercise under the guidance and the supervision of their carers.

The key elements of ACANTO are a robotic friend (the FriWalk) that sup-ports the user in the execution of daily activities that require physical exercise and an intelligent system that recommends activites that a senior user perceives as compelling and rewarding.

The FriWalk takes the form of a standard walking assistant, but it is in fact an intelligent robot that is able to localise itself, to sense the surrounding environment, to plan a course of action that suits the user needs and to guide the user along safe routes. The FriWalk is also a personal trainer that can support the user in the execution of a training programme, monitor the motion of the user in search of muscular or gait problems and report them into the user profile (that can be inspected by doctors and physicians).

The second key idea of ACANTO is that physical exercise is actually "concealed" within compelling activities (such as shopping, taking walks in museums and exhibitions etc.). Such activities have a social dimension (they are proposed to group of users) and are chosen based on the interest and on the past experiences of the user. At the heart of the recommendation system there is a social network which is created and developed by primarily using information collected by the FriWalk using "physical" observations on her behaviour and on her emotional state. For this reason, we call this social network "cyberphysical".

This project aims at developping an autonomous system to drive groups of citizens with respect to point of interest. Those citizens are supposed to communicate, and one of the objective of Tamis is to build a robust and secure system to guarantee this communication. Axel Legay and Olivier Zendra are the permanent researchers of Tamis involved in this project. The project supports two postdocs in Tamis.

9.3.1.2. DIVIDEND

Participant: Laurent Morin.

Start: 2014. End: 2017.

The DIVIDEND project (http://www.chistera.eu/projects/dividend) attacks the data centre energy efficiency bottleneck through vertical integration, specialisation, and cross-layer optimization. Our vision is to present heterogeneous data centres, combining CPUs, GPUs, and task-specific accelerators, as a unified entity to the application developer and let the runtime optimize the utilization of the system resources during task execution. DIVIDEND embraces heterogeneity to dramatically lower the energy per task through extensive hardware specialisation while maintaining the ease of programmability of a homogeneous architecture. To lower communication latency and energy, DIVIDEND refers a lean point-to-point messaging fabric over complex connection-oriented network protocols. DIVIDEND addresses the programmability challenge by adapting and extending the industry-led heterogeneous systems architecture programming language and runtime initiative to account for energy awareness and data movement. DIVIDEND provides for a cross-layer energy optimization framework via a set of APIs for energy accounting and feedback between hardware, compilation, runtime, and application layers. The DIVIDEND project will usher in a new class of vertically integrated data centres and will take a first stab at resolving the energy crisis by improving the power usage effectiveness of data centres.

Laurent Morin from Tamis is involved in this project

9.3.1.3. EMC^2

Participants: Axel Legay, Olivier Zendra.

Start: 2014. End: 2017.

EMC^2 (Embedded Multi-Core systems for Mixed Criticality applications in dynamic and changeable real-time environments https://www.artemis-emc2.eu) is an ARTEMIS Joint Undertaking project in the Innovation Pilot Programme 'Computing platforms for embedded systems' (AIPP5). Embedded systems are the key innovation driver to improve almost all mechatronic products with cheaper and even new functionalities. They support today's information society as inter-system communication enabler. A major industrial challenge arises from the need to face cost efficient integration of different applications with different levels of safety and security on a single computing platform in an open context. EMC^2 finds solutions for dynamic adaptability in open systems, provides handling of mixed criticality applications under real-time conditions, scalability and utmost flexibility, full scale deployment and management of integrated tool chains, through the entire lifecycle. The objective of EMC^2 is to establish Multi-Core technology in all relevant Embedded Systems domains. EMC^2 is a project of 99 partners of embedded industry and research from 19 European countries with an effort of about 800 person years and a total budget of about 100 million Euro.

EMC2 (2014–2017) is at the border between formal methods and security. We in Tamis are mainly using the fundings to develop the Plasma toolset that is used by our statistical model checking and symbolic model checking tools. The permanent members of Tamis who are involved are Axel Legay and Olivier Zendra. The project was initiated during the lifetime of the ESTASYS.Inria team.

9.3.1.4. ENABLE-S3

Participants: Axel Legay, Jean-Louis Lanet.

Start: 2016. End: 2019.

The objective of ENABLE-S3 (http://www.enable-s3.eu) is to establish cost-efficient cross-domain virtual and semi-virtual V&V platforms and methods for ACPS. Advanced functional, safety and security test methods will be developed in order to significantly reduce the verification and validation time but preserve the validity of the tests for the requested high operation range. ENABLE-S3 aspires to substitute today's physical validation and verification efforts by virtual testing and verification, coverage-oriented test selection methods and standardization. ENABLE-S3 is use-case driven; these use cases represent relevant environments and scenarios. Each of the models, methods and tools integrated into the validation platform will be applied to at least one use case (under the guidance of the V&V methodology), where they will be validated (TRL 5) and their usability demonstrated (TRL6). Representative use cases and according applications provide the base for the requirements of methods and tools, as well as for the evaluation of automated systems and respective safety.

This project is industry driven and has the objective of designing new technologies for autonomous transportation, including to secure them. Tamis tests its results on the case studies of the project. Axel Legay and Jean-Louis Lanet are involved in this project. The project supports one postdoc in Tamis starting in 2017.

9.4. International Research Visitors

9.4.1. Visits of International Scientists

- Clémentine MAURICE (Graz University of Technology, Institute of Applied Information Processing and Communications, Austria) visited Tamis and also gave a talk on Reverse-engineering CPUs for fun and profit.

9.4.2. Visits to International Teams

- Axel Legay stayed at Namur University, Belgium.
- Axel Legay stayed at University of Limerick, Ireland.
- Axel Legay and Sean Sedwards stayed at Aalborg University, Denmark.
- Axel Legay, Fabrizio Biondi and Thomas Given-Wilson stayed at John Hopkins University, USA.

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events Organisation

10.1.1.1. General Chair, Scientific Chair

- Axel Legay has been the general chair for the 11th International Conference on Risks and Security of Internet and Systems

10.1.1.2. Member of Organizing Committees

- Axel Legay has been organizing the ICT-Energy Science Conference 2016.

10.1.2. Scientific Events Selection

10.1.2.1. Member of Conference Steering Committees

- Olivier Zendra is a founder and a member of the Steering Committee of ICOOOLPS (International Workshop on Implementation, Compilation, Optimization of OO Languages, Programs and Systems)

10.1.2.2. Chair of Conference Program Committees

- Axel Legay has been the chair for the 14'International Symposium on Automated Technology for Verification and Analysis

10.1.2.3. Member of Conference Program Committees

- Axel Legay has been PC member for ASE, MEMOCODE, FASE, RV, SPLC, FORMATS, FORMALIZE, SETTA,
- Jean-Louis Lanet has been PC member of Cardis 2016, 15th Smart Card Research and Advanced Application Conference, Crisis 2016 The Eleventh International Conference on Risks and Security of Internet and Systems CRiSIS 2016, GraMSec'16, The Third International Workshop on Graphical Models for Security, June 27th Lisbon, Portugal Ressi 2016, Rendez-Vous de la Recherche et de l'Enseignement de la Sécurité des Systèmes d'Information, Toulouse, France, Afadl2016, 15èmes Journées Francophones Internationales sur les Approches Formelles dans l'Assistance au Développement de Logiciels.
- Olivier Zendra has been PC member of PEC 2016 (International Conference on Pervasive and Embedded Computing)

10.1.3. Journal

10.1.3.1. Member of the Editorial Boards

- Axel Legay is a funder and member of the editorial board of "Foundations for Mastering Changes" journal.

10.1.3.2. Reviewer - Reviewing Activities

- Axel Legay has been reviewer for TCS, TSE, Information and Computation.

10.1.4. Invited Talks

- Axel Legay has been an invited speaker for the 10th International Workshop on Reachability Problems.
- Axel Legay has been an invited speaker for the ICT-Energy Science Conference 2016.
- The Wheel of Fault Injection, J.-L Lanet, Workshop Sertif, Grenoble, October 2016.
- Christian Grothoff. "Enabling Secure Web Payments with GNU Taler". Keynote at SPACE 2016 (December).
- Christian Grothoff. "Anonymous Payment Systems" at MAPPING Second General Assembly, Prague, 2016.
- Christian Grothoff. "Netzwerksicherheit: Probleme und Lösungsansätze" at NPO Kongress, Wien, 2016.
- Christian Grothoff. "The GNU Name System: A Public Key Infrastructure for Social Movements in the Age of Universal Surveillance" at Johns Hopkins University, Baltimore, USA, 2016.
- Christian Grothoff. "GNU Taler" at the Free Software Foundation Fellowship Meeting, Düsseldorf, 2016.
- Christian Grothoff. "The GNU Name System: A clean-slate solution to the DNS security and privacy nightmare" at Journée du Conseil scientifique de l'Afnic, Paris, 2016.
- Christian Grothoff. "GNU Taler: A privacy-preserving online payment system for libre society" at CubaConf, Havana, 2016.
- Jeffrey Burdges. "GNU Taler" at the Internet Freedom Festival, 2016.
- Jeffrey Burdges. Preliminary report "Xolotl A compact mixnet format with stronger forward secrecy and hybrid anonymity" at the GNU Hacker Meeting, 2016.
- Florian Dold presented "GNU Taler – Privacy preserving payments for the web" at the GNU Hacker Meeting, 2016.

- Jeffrey Burdges. Panel on “Privacy-preserving decentralization: what challenges are lying ahead?” at the ECRYPT 2016 Workshop on Strategic Research Challenges for Privacy Technologies.
- Christian Grothoff. Panel on "Innovation, Complexity, Risk and Trust" at MAPPING Second General Assembly, Prague, 2016.

10.1.5. Scientific Expertise

- Axel Legay is an expert for the Wallonie Government.
- Axel Legay is a member of Inria’s evaluation committee. He participated to the CR2 and CR1 juries for Lille Center.
- Axel Legay has been in the jury for the chair on cyber security at CentralSupelec.
- Jeffrey Burdges, Christian Grothoff, and Florian Dold have been involved in the W3C Payments Working Group, primarily contributing security and privacy comments on their evolving standard.
- Olivier Zendra is a CIR expert for the MENESR.
- Olivier Zendra is a member of Inria’s evaluation committee. He participated to the CR2 jury for Grenoble Center, to the national CR1 promotion jury, and to the workgroup on the creation of the PACAP team of Inria Rennes.
- Olivier Zendra is a member of the editorial board and co-author of the “HiPEAC Vision” [69]

10.1.6. Research Administration

- Axel Legay is a member of Inria’s evaluation committee.
- Axel Legay is the Representative for non-permanent staff committees (in charge of postdocs).
- Olivier Zendra is a member of Inria’s evaluation committee.
- Olivier Zendra was a member of Inria’s Parity and Equal Opportunities committee.
- Olivier Zendra is a member of Inria’s worksgroup on Inria’s social barometer.
- Olivier Zendra was a member of Inria’s CNHSCT.
- Olivier Zendra was Head of Inria Nancy’s IES Committee (formerly IST).

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

- Master : Axel Legay, Introduction au Model Checking, 36, M2, Université de Bretagne Sud, France
- Master : Axel Legay, Introduction à l’analyse de risques, M2, Université de Bretagne Sud, France

10.2.2. Supervision

- PhD : Aymerick Savary, De la génération de suites de test à partir de modèles formels, University of Sherbrook and University of Limoges, 30th June 2016, Marc Frappier, Jean-Louis Lanet
- PhD : Tiana Razafindralambo, Attaques combinées sur appareil mobiles, University of Limoges, November 2016, Christophe Clavier, Jean-Louis Lanet
- PhD : Neal Walfield, Location prediction for context-aware applications, Johns Hopkins, 4th October 2016, Christian Grothoff
- PhD in progress : Kevin Bukasa, Démarrage sécurisé, 2015, Jean-Louis Lanet and Axel Legay
- PhD in progress : Mounir Chadli (Rennes 1), On Scheduling and SMC, December 2014, Axel Legay and Saddek Bensalem.
- PhD in progress : Olivier Descourbe, On Code Obfuscation, October 2016, Axel Legay and Fabrizio Biondi.
- PhD in progress : Mike Enescu, On Symbolic Execution for Malware Detection, October 2016, Axel Legay and Flavio Oquendo and Fabrizio Biondi.

- PhD in progress : Alexandre Gonsalvez, On Obfuscation via crypto primitives, April 2016, Axel Legay and Caroline Fontaine.
- PhD in progress : Nisrine Jafri (Rennes1), On fault Injection detection with MC of Binary code, December 2015, Axel Legay and Jean-Louis Lanet.
- PhD in progress : Razika Lounas, Validation des spécifications formelles de la mise à jour dynamique des applications Java Card, 2010, Mohamed Mezghiche and Jean-Louis Lanet
- PhD in progress : Aurélien Palisse, Observabilité de codes hostiles, 2015, Jean-Louis Lanet
- PhD in progress : Aurélien Trulla, Caractérisation de malware Android par suivi de flux d'information et nouvelles techniques d'évasion, 2016, Valerie Viet Triem Tong and Jean-Louis Lanet
- PhD in progress : Tristan Ninet (Rennes 1), Vérification formelle d'une implémentation de la pile protocolaire IKEv2, December 2016, Axel Legay, Romaric Maillard and Olivier Zendra

10.2.3. Juries

- Axel Legay has been a referee for the PhD defense of Najah Ben Said (University of Grenoble Alpes).
- Axel Legay has been a member of the jury for the PhD defense of Zaruhi Aslanyan (DTU Denmark).
- Jean-Louis Lanet has been a referee for the PhD defense of Pierre Belgarric (Télécom ParisTech).
- Jean-Louis Lanet has been a referee for the PhD defense of Louis Dureuil (University of Grenoble Alpes).
- Jean-Louis Lanet has been a referee for the PhD defense of Gabriel Risterucci (University of Aix Marseille).
- Jean-Louis Lanet has been a member of the jury for the PhD defense of Benoît Morgan (University of Toulouse).
- Jean-Louis Lanet has been a member of the jury for the PhD defense of Najah Ben Said (University of Grenoble Alpes).
- Olivier Zendra has been a co-referee for the PhD defense of Rabah Laouadi (University of Montpellier).

10.3. Popularization

- Vulnerability Prediction Against Fault Attacks , N. Jafri, A. Legay, J.-L. Lanet, Ercim news 106, 2016
- Skyfall : Tombé du ciel, J.-L. Lanet, Interstices, 2016 In this publication we revisit the movie Skyfall and explain on which scientific background rely some elements of the movie.
- FIC 2016 Internet des objets : la nouvelle fragilité ? We have been invited to participate at a panel with layers, IoT designer to discuss the security of the IoT.
- Atlantico, Et si les objets connectés étaient la plus grande faille qu'entreprises et particuliers pouvaient offrir aux hackers ? January 2016. In this interview we explain that the security is not the main concern of low end IoT, which is not the case with high end IoT.

11. Bibliography

Publications of the year

Articles in International Peer-Reviewed Journal

[1] F. BIONDI, T. GIVEN-WILSON, A. LEGAY.*Attainable Unconditional Security for Shared-Key Cryptosystems*, in "Information Sciences", September 2016, https://hal.inria.fr/hal-01378640.

[2] N. JANATI, G. BOUFFARD, J.-L. LANET, S. ELHAJJI.*Trust Can be Misplaced*, in "Journal of Cryptographic Engineering", 2016 [*DOI :* 10.1007/s13389-016-0142-5], https://hal.inria.fr/hal-01405463.

[3] C. JEGOUREL, A. LEGAY, S. SEDWARDS.*Command-based importance sampling for statistical model checking*, in "Theoretical Computer Science", 2016, vol. 649, p. 1 - 24, https://hal.inria.fr/hal-01387299.

[4] J. H. KIM, A. LEGAY, L.-M. TRAONOUEZ, A. BOUDJADAR, U. NYMAN, K. G. LARSEN, I. LEE, J.-Y. CHOI.*Optimizing the Resource Requirements of Hierarchical Scheduling Systems*, in "ACM SIGBED Review", June 2016, vol. 13, n^{o} 3, p. 41 - 48 [*DOI :* 10.1145/2983185.2983192], https://hal.inria.fr/hal-01406459.

[5] T. T. H. LE, R. PASSERONE, U. FAHRENBERG, A. LEGAY.*A Tag Contract Framework for Modeling Heterogeneous Systems*, in "Science of Computer Programming", January 2016, https://hal.inria.fr/hal-01406446.

[6] T. T. H. LE, R. PASSERONE, U. FAHRENBERG, A. LEGAY.*Contract-Based Requirement Modularization via Synthesis of Correct Decompositions*, in "Theory of Computing Systems", May 2016, vol. 15 [*DOI :* 10.1145/2885752], https://hal.inria.fr/hal-01406481.

[7] A. LEGAY, L.-M. TRAONOUEZ.*Statistical Model Checking with Change Detection*, in "Foundations for Mastering Change", September 2016, https://hal.archives-ouvertes.fr/hal-01242138.

[8] F. MOZHDEH, J.-L. LANET.*Chronicle of a Java Card Death*, in "Journal of Computer Virology and Hacking Techniques", 2016 [*DOI :* 10.1007/s11416-016-0276-0], https://hal.inria.fr/hal-01385197.

[9] A. NOURI, M. BOZGA, A. MOLNOS, A. LEGAY, S. BENSALEM.*ASTROLABE: A Rigorous Approach for System-Level Performance Modeling and Analysis*, in "ACM Transactions on Embedded Computing Systems (TECS)", May 2016 [*DOI :* 10.1145/2885498], https://hal.inria.fr/hal-01406474.

Invited Conferences

[10] A. LEGAY, S. SEDWARDS, L.-M. TRAONOUEZ.*Rare Events for Statistical Model Checking: An Overview*, in "Reachability Problems", Aalborg, Denmark, September 2016, vol. 9899, p. 23 - 35, https://hal.inria.fr/hal-01387406.

International Conferences with Proceedings

[11] A. ARNOLD, M. BALEANI, A. FERRARI, M. MARAZZA, V. SENNI, A. LEGAY, J. QUILBEUF, C. ETZIEN.*An Application of SMC to continuous validation of heterogeneous systems*, in "Simutools 2016 - Ninth EAI International Conference on Simulation Tools and Techniques", Prague, Czech Republic, August 2016, https://hal.inria.fr/hal-01390487.

[12] F. BIONDI, A. LEGAY.*Security and Privacy of Protocols and Software with Formal Methods*, in "7th International Symposium on Leveraging Applications of Formal Methods, Verification and Validation", Kommeno, Greece, October 2016, https://hal.inria.fr/hal-01378645.

[13] J. BURDGES, F. DOLD, C. GROTHOFF, M. STANISCI.*Taler: Usable, privacy-preserving payments for the Web*, in "HotPETS 2016 - Workshop on Hot Topics in Privacy Enhancing Technologies", Darmstadt , Germany, June 2016, https://hal.inria.fr/hal-01398201.

[14] E. CAVALCANTE, J. QUILBEUF, L.-M. TRAONOUEZ, F. OQUENDO, T. BATISTA, A. LEGAY.*Statistical Model Checking of Dynamic Software Architectures*, in "ECSA 2016 - 10th European Conference on Software Architecture", Copenhague, Denmark, Proceedings of the 2016 European Conference of Software Architecture, November 2016, https://hal.inria.fr/hal-01390707.

[15] M. CHADLI, J. H. KIM, A. LEGAY, L.-M. TRAONOUEZ, S. NAUJOKAT, B. STEFFEN, K. G. LARSEN.*A Model-Based Framework for the Specification and Analysis of Hierarchical Scheduling Systems*, in "FMICS-AVoCS", Pise, Italy, Critical Systems: Formal Methods and Automated Verification, Springer, September 2016, vol. 9933, p. 133 - 141 [*DOI :* 10.1007/978-3-319-45943-1_9], https://hal.archives-ouvertes.fr/hal-01241681.

[16] D. COUROUSSÉ, T. BARRY, B. ROBISSON, P. JAILLON, O. POTIN, J.-L. LANET.*Runtime Code Polymorphism as a Protection Against Side Channel Attacks*, in "10th IFIP WG 11.2 International Conference, WISTP 2016", Heraklion, Greece, Information Security Theory and Practice, September 2016, vol. 9895, p. 136-152 [*DOI :* 10.1007/978-3-319-45931-8_9], https://hal-emse.ccsd.cnrs.fr/emse-01372223.

[17] P. R. D'ARGENIO, A. HARTMANNS, A. LEGAY, S. SEDWARDS.*Statistical Approximation of Optimal Schedulers for Probabilistic Timed Automata*, in "Integrated Formal Methods", Reykjavik, Iceland, June 2016, vol. 9681, p. 99-114, https://hal.inria.fr/hal-01387362.

[18] X. DEVROEY, G. PERROUIN, A. LEGAY, P.-Y. SCHOBBENS, P. HEYMANS.*Search-based Similarity-driven Behavioural SPL Testing*, in "VaMoS '16 - Tenth International Workshop on Variability Modelling of Software-intensive Systems", Salvador, Brazil, ACM, January 2016, p. 89 - 96 [*DOI :* 10.1145/2866614.2866627], https://hal.inria.fr/hal-01406585.

[19] X. DEVROEY, G. PERROUIN, M. PAPADAKIS, A. LEGAY, P.-Y. SCHOBBENS, P. HEYMANS.*Featured model-based mutation analysis*, in "ICSE '16 - 38th International Conference on Software Engineering", Austin, United States, May 2016, p. 655 - 666 [*DOI :* 10.1145/2884781.2884821], https://hal.inria.fr/hal-01406512.

[20] O. GADYATSKAYA, R. RYDHOF HANSEN, K. G. LARSEN, A. LEGAY, M. C. OLESEN, D. B. POULSEN.*Modelling Attack-defense Trees Using Timed Automata*, in "FORMATS 2016 -14th International Conference on Formal Modelling and Analysis of Timed Systems", Quebec City, Canada, August 2016, p. 35 - 50 [*DOI :* 10.1007/978-3-319-44878-7_3], https://hal.inria.fr/hal-01406706.

[21] Á. GARCÍA-RECUERO, J. BURDGES, C. GROTHOFF.*Privacy-Preserving Abuse Detection in Future Decentralised Online Social Networks*, in "11th International ESORICS Workshop in Data Privacy Management, DPM 2016", Heraklion, Crete, Greece, G. LIVRAGA, V. TORRA, A. ALDINI, F. MARTINELLI, N. SURI (editors), Springer Lecture Notes in Computer Science (LNCS) series, Springer, September 2016, vol. 9963, p. 78-93 [*DOI :* 10.1007/978-3-319-47072-6_6], https://hal.inria.fr/hal-01355951.

[22] T. GIVEN-WILSON, A. LEGAY.*On the Expressiveness of Symmetric Communication*, in "Theoretical Aspects of Computing – ICTAC 2016", Taipei, Taiwan, October 2016, vol. 9965, p. 139-157 [*DOI :* 10.1007/978-3-319-46750-4_9], https://hal.inria.fr/hal-01241839.

[23] A. HEUSER, S. PICEK, S. GUILLEY, N. MENTENS.*Side-channel Analysis of Lightweight Ciphers: Does Lightweight Equal Easy?*, in "RFIDSec 2016: 12th Workshop on RFID and IoT Security", Hong Kong, Hong Kong SAR China, November 2016, https://hal.inria.fr/hal-01402238.

[24] Y. KAWAMOTO, F. BIONDI, A. LEGAY.*Hybrid Statistical Estimation of Mutual Information for Quantifying Information Flow*, in "FM 2016 - 21st International Symposium on Formal Methods", Limassol, Cyprus, November 2016, https://hal.inria.fr/hal-01378675.

[25] J. H. KIM, A. LEGAY, L.-M. TRAONOUEZ, M. ACHER, S. KANG.*A Formal Modeling and Analysis Framework for Software Product Line of Preemptive Real-Time Systems*, in "Symposium on Applied Computing", Pise, Italy, Proceedings of the 31st Annual ACM Symposium on Applied Computing, ACM, April 2016, p. 1562 - 1565 [*DOI :* 10.1145/2851613.2851977], https://hal.archives-ouvertes.fr/hal-01241673.

[26] K. G. LARSEN, A. LEGAY.*On the Power of Statistical Model Checking*, in "7th International Symposium, ISoLA 2016", Corfu, Greece, October 2016, p. 843 - 862 [*DOI :* 10.1007/978-3-319-47169-3_62], https://hal.inria.fr/hal-01406537.

[27] H. LE BOUDER, T. BARRY, D. COUROUSSÉ, J.-L. LANET, R. LASHERMES.*A Template Attack Against VERIFY PIN Algorithms*, in "SECRYPT 2016", Lsbonne, Portugal, July 2016, p. 231 - 238 [*DOI :* 10.5220/0005955102310238], https://hal.inria.fr/hal-01383143.

[28] H. LE BOUDER, R. LASHERMES, Y. LINGE, G. THOMAS, J.-Y. ZIE.*A Multi-Round Side Channel Attack on AES using Belief Propagation*, in "FPS 2016", Québec, Canada, October 2016, https://hal.inria.fr/hal-01405793.

[29] A. LEGAY, S. SEDWARDS, L.-M. TRAONOUEZ.*Plasma Lab: A Modular Statistical Model Checking Platform*, in "ISoLA", Corfou, Greece, Leveraging Applications of Formal Methods, Verification and Validation: Foundational Techniques, Springer, October 2016, vol. 9952, p. 77 - 93 [*DOI :* 10.1007/978-3-319-47166-2_6], https://hal.inria.fr/hal-01387435.

[30] R. LOUNAS, N. JAFRI, A. LEGAY, M. MEZGHICHE, J.-L. LANET.*A Formal Verification of Safe Update Point Detection in Dynamic Software Updating*, in "International Conference on Risks and Security of Internet and Systems", Roscoff, France, LNCS, Telecom Bretagne, September 2016, https://hal.inria.fr/hal-01405467.

[31] R. LOUNAS, N. JAFRI, A. LEGAY, M. MEZGHICHE, J.-L. LANET.*A Formal Verification of Safe Update Point Detection in Dynamic Software Updating*, in "The 11th International Conference on Risks and Security of Internet and Systems (CRiSIS 2016)", Roscoff, France, September 2016, https://hal.inria.fr/hal-01400051.

[32] A. MESBAH, L. REGNAUD, J.-L. LANET, M. MEZGHICHE.*The Hell Forgery, Polymorphic Codes Shoot Again*, in "15th Smart Card Research and Advanced Application Conference", Cannes, France, Aurélien Francillon, November 2016, https://hal.inria.fr/hal-01385202.

[33] F. MOZHDEH, J.-L. LANET.*Paper Tiger, an endless fight*, in "SecITC 9th International Conference on Security for Information Technology and Communications", Bucarest, Romania, Innovative Security Solutions for Information Technology and Communications, June 2016, vol. LNCS, n° 10006 [*DOI :* 10.1007/978-3-319-47238-6], https://hal.inria.fr/hal-01385201.

[34] A. MRABET, N. EL-MRABET, R. LASHERMES, J.-B. RIGAUD, B. BOUALLEGUE, S. MESNAGER, M. MACHHOUT.*High-performance Elliptic Curve Cryptography by Using the CIOS Method for Modular Multiplication*, in "CRiSIS 2016", Roscoff, France, September 2016, https://hal.inria.fr/hal-01383162.

[35] V. C. NGO, A. LEGAY, V. JOLOBOFF.*PSCV: A Runtime Verification Tool for Probabilistic SystemC Models*, in "CAV 2016 - 28th International Conference on Computer Aided Verification", Toronto, Canada, S.

CHAUDHURI, A. FARZAN (editors), LNCS - Lecture Notes in Computer Science, Springer, July 2016, vol. 9779, p. 84 - 91 [*DOI :* 10.1007/978-3-319-41528-4_5], https://hal.inria.fr/hal-01406488.

[36] A. NOURI, M. BOZGA, A. LEGAY, S. BENSALEM.*Performance Evaluation of Complex Systems Using the SBIP Framework*, in "he International Conference on Verification and Evaluation of Computer and Communication Systems (VECoS)", Tunis, Tunisia, October 2016, https://hal.inria.fr/hal-01406591.

[37] R. OLAECHEA, U. FAHRENBERG, J. M. ATLEE, A. LEGAY.*Long-term average cost in featured transition systems*, in "SPLC '16 - 20th International Systems and Software Product Line Conference", Beijing, China, ACM, September 2016, p. 109 - 118 [*DOI :* 10.1145/2934466.2934473], https://hal.inria.fr/hal-01406541.

[38] R. OLAECHEA, U. FAHRENBERG, M. JOANNE, A. LEGAY.*Long-Term Average Cost in Featured Transition Systems*, in "the 20th International Systems and Software Product Line Conference", Beijing, China, September 2016, https://hal.inria.fr/hal-01406606.

[39] F. OQUENDO, J. BUISSON, E. LEROUX, G. MOGUÉROU, J. QUILBEUF.*The SoS Architect Studio: Toolchain for the Formal Architecture Description and Analysis of Software-intensive Systems-of-Systems with SosADL*, in "Proceedings of the ECSA International Colloquium on Software-intensive Systems-of-Systems (SiSoS)", Copenhagen, Denmark, November 2016, https://hal.archives-ouvertes.fr/hal-01443130.

[40] G. PERROUIN, M. AMRANI, M. ACHER, B. COMBEMALE, A. LEGAY, P.-Y. SCHOBBENS.*Featured model types: Towards Systematic Reuse in Modelling Language Engineering*, in "MiSE '16 - 8th International Workshop on Modeling in Software Engineering", New york, United States, ACM, May 2016, p. 1 - 7 [*DOI :* 10.1145/2896982.2896987], https://hal.inria.fr/hal-01406507.

[41] J. QUILBEUF, E. CAVALCANTE, L.-M. TRAONOUEZ, F. OQUENDO, T. BATISTA, A. LEGAY.*A Logic for the Statistical Model Checking of Dynamic Software Architectures*, in "ISoLA", Corfou, Greece, Leveraging Applications of Formal Methods, Verification and Validation: Foundational Techniques, Springer, October 2016, vol. 9952, p. 806 - 820 [*DOI :* 10.1007/978-3-319-47166-2_56], https://hal.inria.fr/hal-01387429.

[42] M. H. TER BEEK, A. LEGAY, A. LLUCH LAFUENTE, A. VANDIN.*Statistical Model Checking for Product Lines* , in "7th International Symposium, ISoLA 2016", Corfu, Greece, October 2016, https://hal.inria.fr/hal-01406531.

[43] N. H. WALFIELD, J. L. GRIFFIN, C. GROTHOFF.*A Quantitative Analysis of Cell Tower Trace Data for Understanding Human Mobility and Mobile Networks*, in "6th International Workshop on Mobile Entity Localization, Tracking and Analysis (MELT)", San Francisco, United States, October 2016 [*DOI :* 10.1145/1235], https://hal.inria.fr/hal-01378622.

Conferences without Proceedings

[44] M. BOIZARD, J.-L. LANET.*Internet des objets : la nouvelle fragilité ? : Atelier*, in "Forum international de la cybersécurité", Lille, France, Région Nord-Pas de Calais, Euratechnologies, la Gendarmerie nationale, CEIS, January 2016, https://halshs.archives-ouvertes.fr/halshs-01259237.

[45] D. COUROUSSÉ, J.-L. LANET, B. ROBISSON, T. BARRY, P. JAILLON.*COGITO: Génération de code au runtime pour la sécurité des systèmes embarqué*, in "Ressi 2016", Toulouse, France, May 2016, https://hal.inria.fr/hal-01405776.

[46] J. DUCHÊNE, C. LE GUERNIC, E. ALATA, V. NICOMETTE, M. KAÂNICHE.*Protocol reverse engineering: Challenges and obfuscation*, in "The 11th International Conference on Risks and Security of Internet and Systems", Roscoff, France, September 2016, https://hal.inria.fr/hal-01388060.

[47] M. GRAA, F. CUPPENS, N. CUPPENS-BOULAHIA, J.-L. LANET.*Tracking explicit and control flows in Java and native Android apps code,*, in "ICISSP 2016 : 2nd International Conference on Information Systems Security and Privacy,", Roma, Italy, February 2016, https://hal.inria.fr/hal-01385196.

[48] Á. GARCÍA-RECUERO.*Discouraging Abusive Behavior in Privacy-Preserving Online Social Networking Applications*, in "In 25th International World Wide Web Conference", Montréal, Canada, April 2016, vol. WWW Companion Volume, https://hal.inria.fr/hal-01272654.

[49] C. JEGOUREL, K. G. LARSEN, A. LEGAY, M. MIKUČIONIS, D. B. POULSEN, S. SEDWARDS.*Importance Sampling for Stochastic Timed Automata*, in "Dependable Software Engineering: Theories, Tools, and Applications", Beijing, China, November 2016, https://hal.inria.fr/hal-01387293.

[50] V. C. NGO, A. LEGAY, J. QUILBEUF.*Statistical Model Checking for SystemC Models*, in "High Assurance Systems Engineering Symposium", Orlando, Florida, United States, January 2016, https://hal.inria.fr/hal-01238162.

[51] A. PALISSE, A. DURAND, J.-L. LANET.*Malware'O'Matic a platform to analyze Malware*, in "French Japanese workshop on CyberSecurity", Rennes, France, Inria, September 2016, https://hal.inria.fr/hal-01405636.

[52] A. PALISSE, J.-L. LANET.*Analyse et détection de logiciels malveillants,* , in "Ressi 2016", Toulouse, France, May 2016, https://hal.inria.fr/hal-01405627.

[53] A. PALISSE, H. LE BOUDER, C. LE GUERNIC, A. LEGAY, J.-L. LANET.*Ransomware and the Legacy Crypto API*, in "The 11th International Conference on Risks and Security of Internet and Systems", Roscoff, France, September 2016, https://hal.inria.fr/hal-01388056.

Scientific Books (or Scientific Book chapters)

[54] S. CHAUMETTE, J. H. KIM, K. NAMUDURI, J. P. STERBENZ (editors). *UAV Networks and Communications*, Cambridge University Press, 2016, https://hal.archives-ouvertes.fr/hal-01391820.

Books or Proceedings Editing

[55] F. OQUENDO, M. ALI BABAR, K. DRIRA, A. LEGAY (editors). *Proceedings of the European Colloquium on Software-intensive Systems-of-Systems (ECSoS 2016)*, ACM DL, Copenhagen, Denmark, December 2016, https://hal.archives-ouvertes.fr/hal-01445339.

[56] F. OQUENDO, K. DRIRA, A. LEGAY, T. BATISTA (editors). *Proceedings of the 1st ACM SAC Conference Track on Software-intensive Systems-of-Systems (SiSoS 2017): 32nd ACM SIGAPP Symposium On Applied Computing*, ACM, Marrakesh, Morocco, April 2017, https://hal.archives-ouvertes.fr/hal-01445350.

Scientific Popularization

[57] F. BIONDI, S. JOSSE, A. LEGAY.*Bypassing Malware Obfuscation with Dynamic Synthesis*, in "ERCIM News", September 2016, n^{o} 106, https://hal.inria.fr/hal-01378662.

[58] C. GROTHOFF, J. PORUP.*The NSA's SKYNET program may be killing thousands of innocent people*, in "Ars Technica", February 2016, https://hal.inria.fr/hal-01278193.

[59] N. JAFRI, A. LEGAY, J.-L. LANET.*Vulnerability Prediction Against Fault Attacks*, in "ERCIM News", July 2016, https://hal.inria.fr/hal-01394973.

[60] J.-L. LANET.*Skyfall : Tombé du ciel*, in "Interstices", June 2016, https://hal.inria.fr/hal-01352807.

[61] H. LE BOUDER.*Des attaques informatiques utilisant la physique*, in "Interstices", November 2016, https://hal.inria.fr/hal-01408630.

Other Publications

[62] F. BIONDI, M. CHADLI, T. GIVEN-WILSON, A. LEGAY.*Information Leakage as a Scheduling Resource*, October 2016, working paper or preprint, https://hal.inria.fr/hal-01382052.

[63] F. BIONDI, T. GIVEN-WILSON, A. LEGAY.*Attainable Unconditional Security for Shared-Key Cryptosystems*, April 2016, working paper or preprint, https://hal.inria.fr/hal-01233185.

[64] Q. CAPPART, C. LIMBRÉE, P. SCHAUS, J. QUILBEUF, L.-M. TRAONOUEZ, A. LEGAY.*Verification of interlocking systems using statistical model checking*, November 2016, working paper or preprint, https://hal.inria.fr/hal-01398649.

[65] U. FAHRENBERG, A. LEGAY.*A Linear-Time–Branching-Time Spectrum of Behavioral Specification Theories*, October 2016, We propose behavioral specification theories for most equivalences in the linear-time–branching-time spectrum, https://hal.inria.fr/hal-01406603.

[66] T. GIVEN-WILSON, N. JAFRI, J.-L. LANET, A. LEGAY.*An Automated Formal Process for Detecting Fault Injection Vulnerabilities in Binaries*, January 2017, working paper or preprint, https://hal.inria.fr/hal-01400283.

[67] A. HEUSER, S. PICEK, S. GUILLEY, N. MENTENS.*Side-channel Analysis of Lightweight Ciphers: Current Status and Future Directions*, October 2016, Lightweight Cryptography Workshop 2016, https://hal.inria.fr/hal-01407264.

[68] Y. KAWAMOTO, F. BIONDI, A. LEGAY.*Hybrid Statistical Estimation of Mutual Information for Quantifying Information Flow*, September 2016, working paper or preprint, https://hal.inria.fr/hal-01241360.

References in notes

[69] M. DURANTON, K. DE BOSSCHERE, C. GAMRAT, J. MAEBE, H. MUNK, O. ZENDRA.*HiPEAC (High Performance and Embedded Architecture and Compilation) Vision 2017*, December 2016, p. 1–137, working document, to be published early 2017, https://www.hipeac.net/v17.

[70] A. LEGAY, L.-M. TRAONOUEZ.*Statistical Model Checking of Simulink Models with Plasma Lab*, in " Fourth International Workshop on Formal Techniques for Safety-Critical Systems", Paris, France, November 2015, https://hal.archives-ouvertes.fr/hal-01241249.

[71] A. SAVARY, M. FRAPPIER, M. LEUSCHEL, J. LANET.*Model-Based Robustness Testing in Event-B Using Mutation*, in "Software Engineering and Formal Methods - 13th International Conference, SEFM 2015, York, UK, September 7-11, 2015. Proceedings", R. CALINESCU, B. RUMPE (editors), Lecture Notes in Computer Science, Springer, 2015, vol. 9276, p. 132–147, http://dx.doi.org/10.1007/978-3-319-22969-0_10.

Project-Team TASC

Theory, Algorithms and Systems for Constraints

IN COLLABORATION WITH: Laboratoire d'Informatique de Nantes Atlantique (LINA)

IN PARTNERSHIP WITH:
Ecole des Mines de Nantes

Université Nantes

RESEARCH CENTER
Rennes - Bretagne-Atlantique

THEME
Architecture, Languages and Compilation

Table of contents

Project-Team TASC

Creation of the Project-Team: 2011 January 01, end of the Project-Team: 2016 December 31

Keywords:

Computer Science and Digital Science:

2.1.5. - Constraint programming
3.2.1. - Knowledge bases
3.2.3. - Inference
6.1. - Mathematical Modeling
7.2. - Discrete mathematics, combinatorics
8.2. - Machine learning
8.6. - Decision support

Other Research Topics and Application Domains:

4.1. - Fossile energy production (oil, gas)
4.3. - Renewable energy production
8.3. - Urbanism and urban planning

1. Members

Faculty Members

Nicolas Beldiceanu [Team leader, EMN, Professor, HDR]
Mohamed Amine Balafrej [EMN, Associate Professor, until Sep 2016]
Frédéric Benhamou [Univ. Nantes, Professor]
Xavier Lorca [EMN, Associate Professor, HDR]
Eric Monfroy [Univ. Nantes, Professor, HDR]
Thierry Petit [EMN, Associate Professor, HDR]
Florian Richoux [Univ. Nantes, Associate Professor]
Gilles Simonin [EMN, Associate Professor, from Sep 2016]
Charlotte Truchet [Univ. Nantes, Associate Professor]

Technical Staff

Charles Prud'Homme [ARMINES]

PhD Students

Ekaterina Arafailova [ARMINES]
Anicet Bart [Univ. Nantes]
Gilles Madi Wamba [EMN]
Alejandro Reyes Amaro [Univ. Nantes, until Sep 2016]
Ignacio Salas Donoso [EMN, until Apr 2016]

Post-Doctoral Fellow

Abbas El Dor [ARMINES]

Administrative Assistant

Anne Claire Binetruy [Inria]

2. Overall Objectives

2.1. Objectives of the team

2.1.1. Origin and Current Situation

Constraint programming emerges in the eighties and develops at **the intersection of Artificial Intelligence and Operations Research**, of Computer Science and Mathematics. Multidisciplinary by nature it keeps on using knowledge from various topics such as discrete mathematics, theoretical computer science (graph theory, combinatorics, algorithmic, complexity), functional analysis and optimization, IT and software engineering. Constraint programming was identified in 1996 by the ACM as a *strategic topic for Computer Science*. The turn of the century has seen the development of optimization technology in the industry (with notably Ilog, IBM, Dash and more recently Microsoft, Google and Dynadec) and the corresponding scientific field, at the border of Constraint Programming, Mathematical Programming, Local Search and Numerical Analysis. Optimisation technology is now assisting public sector, companies and people to some extent for making decisions that use resources better and match specific requirements in an increasingly complex world. Indeed, computer aided decision and optimization is becoming one of the cornerstones for providing assistance to all kinds of human activities.

Today, with the preeminence of optimization technology in most industrial sectors, we argue that quick and ad hoc solutions, often used today, cannot support the long-term development of optimization technology and its broad diffusion. We also argue that there should be a much more direct link between mathematical results and their systematic reuse in the main fields of optimization technology.

2.1.2. General Challenges

In spite of its importance, computer aided decision and optimization suffers from a number of fundamental weaknesses that prevent from taking advantage of its full potential and hinder its progress and its capacity to deal with more and more complex situations. This can be mostly blamed on the diversity of actors, which are:

- Spread out in distinct scientific communities, each with its own focus:
  - On the one hand, computer science for providing languages, modelling tools and libraries. While focusing on providing flexible and powerful programming paradigm that can be easily deployed and maintained on modern architectures, it does not address the central question of how to come up in a systematic way with efficient methods for optimization and decision problems.
  - On the other hand, applied mathematics for the theory part. The focus is to come up with powerful abstractions that allow understanding the structure of a class of problems, independently of its practical and systematic uses in modern software components.
- Spread out in distinct technological communities, each independently pushing its own solving paradigm like constraint programming, linear and integer programming, continuous optimization, constraint-based local search (e.g., COMET). To some extent, most of these techniques exploit in different ways the same mathematical results, that are manually adapted to fit the main way to proceed of a given technology.

Thus, a first challenge encountered by constraint programming is the design of computer systems implementing **in a transparent way** effective solving techniques.

- Ideally, the user must be able to **describe his problem in a high level modelling language** without being concerned with the underlying solving mechanisms used. Such systems must also be independent both from any computer programming language and from any resolution engine.
- In order to assists user, systems must also offer **digital knowledge base in problem solving** that make available state of the art models and heuristics for large set of well identified problems.
- Lastly, the user must have the ability to interpret the returned solutions, in particular within the context of **over constrained problems where it is necessary to partly relax some constraints**, and that in the most realistic possible way.

A second challenge resides in the **speed of resolution especially in the context of large-scale data**. One has to adapt techniques such as generic consistency algorithms, graph algorithms, mathematical programming, meta-heuristics and to integrate them within the framework of constraint programming. This integration generates new questions such as the design of incremental algorithms, the automatic decomposition or the automatic reformulation of problems.

Finally a third challenge deals with the use of constraint programming in the context of **complex industrial problems**, especially when both discrete and continuous aspects are present. Complexity has multiple causes such as:

- the combination of temporal and spatial aspects, of continuous and discrete aspects,
- the dynamic character of some phenomena inducing a modification of the constraints and data during time,
- the difficulty of expressing some physical constraints, e.g. load balancing and temporal stability,
- the necessary decomposition of large problems inducing significant solution performance losses.

3. Research Program

3.1. Overview

Basic research is guided by the challenges raised before: to classify and enrich the models, to automate reformulation and resolution, to dissociate declarative and procedural knowledge, to come up with theories and tools that can handle problems involving both continuous and discrete variables, to develop modelling tools and to come up with solving tools that scale well. On the one hand, **classification aspects** of this research are integrated within a knowledge base about combinatorial problem solving: the global constraint catalog (see http://sofdem.github.io/gccat/). On the other hand, **solving aspects** are capitalized within the constraint solving system CHOCO. Lastly, within the framework of its activities of valorisation, teaching and of partnership research, the team uses constraint programming for solving various concrete problems. The challenge is, on one side to increase the visibility of the constraints in the others disciplines of computer science, and on the other side to contribute to a broader diffusion of the constraint programming in the industry.

3.2. Fundamental Research Topics

This part presents the research topics investigated by the project:

- Global Constraints Classification, Reformulation and Filtering,
- Convergence between Discrete and Continuous,
- Dynamic, Interactive and over Constrained Problems,
- Solvers.

These research topics are in fact not independent. The work of the team thus frequently relates transverse aspects such as explained global constraints, Benders decomposition and explanations, flexible and dynamic constraints, linear models and relaxations of constraints.

3.2.1. Constraints Classification, Reformulation and Filtering

In this context our research is focused (a) first on identifying recurring combinatorial structures that can be used for modelling a large variety of optimization problems, and (b) exploit these combinatorial structures in order to come up with efficient algorithms in the different fields of optimization technology. The key idea for achieving point (b) is that many filtering algorithms both in the context of Constraint Programming, Mathematical Programming and Local Search can be interpreted as the maintenance of invariants on specific domains (e.g., graph, geometry). The systematic classification of global constraints and of their relaxation brings a synthetic view of the field. It establishes links between the properties of the concepts used to describe constraints and the properties of the constraints themselves. Together with SICS, the team develops and maintains *a catalog of global constraints*, which describes the semantics of more than 431 constraints, and proposes a unified mathematical model for expressing them. This model is based on graphs, automata and logic formulae and allows to derive filtering methods and automatic reformulation for each constraint in a unified way (see http://www.emn.fr/x-info/sdemasse/gccat/index.html). We consider hybrid methods (i.e., methods that involve more than one optimization technology such as constraint programming, mathematical programming or local search), to draw benefit from the respective advantages of the combined approaches. More fundamentally, the study of hybrid methods makes it possible to compare and connect strategies of resolution specific to each approach for then conceiving new strategies. Beside the works on classical, complete resolution techniques, we also investigate local search techniques from a mathematical point of view. These partly random algorithms have been proven very efficient in practice, although we have little theoretical knowledge on their behaviour, which often makes them problem-specific. Our research in that area is focused on a probabilistic model of local search techniques, from which we want to derive quantified information on their behaviour, in order to use this information directly when designing the algorithms and exploit their performances better. We also consider algorithms that maintain local and global consistencies, for more specific models. Having in mind the trade off between genericity and effectiveness, the effort is put on the efficiency of the algorithms with guarantee on the produced levels of filtering. This effort results in adapting existing techniques of resolution such as graph algorithms. For this purpose we identify necessary conditions of feasibility that can be evaluated by efficient incremental algorithms. Genericity is not neglected in these approaches: on the one hand the constraints we focus on are applicable in many contexts (for example, graph partitioning constraints can be used both in logistics and in phylogeny); on the other hand, this work led to study the portability of such constraints and their independence with specific solvers. This research orientation gathers various work such as strong local consistencies, graph partitioning constraints, geometrical constraints, and optimization and soft constraints. Within the perspective to deal with complex industrial problems, we currently develop meta constraints (e.g. *geost*) handling all together the issues of large-scale problems, dynamic constraints, combination of spatial and temporal dimensions, expression of business rules described with first order logic.

3.2.2. Convergence between Discrete and Continuous

Many industrial problems mix continuous and discrete aspects that respectively correspond to physical (e.g., the position, the speed of an object) and logical (e.g., the identifier, the nature of an object) elements. Typical examples of problems are for instance:

- *Geometrical placement problems* where one has to place in space a set of objects subject to various geometrical constraints (i.e., non-overlapping, distance). In this context, even if the positions of the objects are continuous, the structure of optimal configurations has a discrete nature.
- *Trajectory and mission planning problems* where one has to plan and synchronize the moves of several teams in order to achieve some common goal (i.e., fire fighting, coordination of search in the context of rescue missions, surveillance missions of restricted or large areas).
- *Localization problems in mobile robotic* where a robot has to plan alone (only with its own sensors) its trajectory. This kind of problematic occurs in situations where the GPS cannot be used (e.g., under water or Mars exploration) or when it is not precise enough (e.g., indoor surveillance, observation of contaminated sites).

Beside numerical constraints that mix continuous and integer variables we also have global constraints that involve both type of variables. They typically correspond to graph problems (i.e., graph colouring, domination in a graph) where a graph is dynamically constructed with respect to geometrical and-or temporal constraints. In this context, the key challenge is avoiding decomposing the problem in a discrete and continuous parts as it is traditionally the case. As an illustrative example consider *the wireless network deployment problem*. On the one hand, the continuous part consists of finding out where to place a set of antenna subject to various geometrical constraints. On the other hand, by building an interference graph from the positions of the antenna, the discrete part consists of allocating frequencies to antenna in order to avoid interference. In the context of convergence between discrete and continuous variables, our goals are:

- First to identify and compare typical class of techniques that are used in the context of continuous and discrete solvers.

- To see how one can unify and/or generalize these techniques in order to handle in an integrated way continuous and discrete constraints within the same framework.

3.2.3. Dynamic, Interactive and over Constrained Problems

Some industrial applications are defined by a set of constraints which may change over time, for instance due to an interaction with the user. Many other industrial applications are over-constrained, that is, they are defined by set of constraints which are more or less important and cannot be all satisfied at the same time. Generic, dedicated and explanation-based techniques can be used to deal efficiently with such applications. Especially, these applications rely on the notion of *soft constraints* that are allowed to be (partially) violated. The generic concept that captures a wide variety of soft constraints is the violation measure, which is coupled with specific resolution techniques. Lastly, soft constraints allow to combine the expressive power of global constraints with local search frameworks.

3.2.4. Solvers

- *Discrete solver* Our theoretical work is systematically validated by concrete experimentations. We have in particular for that purpose the CHOCO constraint platform. The team develops and maintains CHOCO initially with the assistance of the laboratory e-lab of Bouygues (G. Rochart), the company Amadeus (F. Laburthe), and others researchers such as N. Jussien and H. Cambazard (4C, INP Grenoble). Since 2008 the main developments are done by Charles Prud'homme and Xavier Lorca. The functionalities of CHOCO are gradually extended with the outcomes of our works: design of constraints, analysis and visualization of explanations, etc. The open source CHOCO library is downloaded on average 450 times each month since 2006. CHOCO is developed in line with the research direction of the team, in an open-minded scientific spirit. Contrarily to other solvers where the efficiency often relies on problem-specific algorithms, CHOCO aims at providing the users both with reusable techniques (based on an up-to-date implementation of the global constraint catalogue) and with a variety of tools to ease the use of these techniques (clear separation between model and resolution, event-based solver, management of the over-constrained problems, explanations, etc.).
- *Discrete continuous* We use discrete convexity to describe filtering for families of constraints: We introduce a propagator for pairs of Sum constraints, where the expressions in the sums respect a form of convexity. This propagator is parametric and can be instantiated for various concrete pairs, including Deviation, Spread, and the conjunction of Linear$\leq$ and Among. We show that despite its generality, our propagator is competitive in theory and practice with state-of-the-art propagators.
- *Constraint programming and verification* Constraint Programming has already had several applications to verification problems. It also has many common ideas with Abstract Interpretation, a theory of approximation of the semantics of programs. In both cases, we are interested in a particular set (solutions in CP, program traces in semantics), which is hard or impossible to compute, and this set is replaced by an over-approximation (consistent domains / abstract domains). Previous works (internship of Julie Laniau, PhD of Marie Pelleau, collaboration with the Abstract Interpretation team at the ENS and Antoine Miné in particular) have exhibited some of these links, and identified some situations where the two fields, Abstract Interpretation and Constraint Programming, can complement each other. It is the case in real-time stream processing languages, where Abstract Interpretation techniques may not be precise enough when analyzing loops. With the PhD of Anicet Bart, we are currently working on using CP techniques to find loop invariants for the Faust language, a functional sound processing language.

This work around the design and the development of solvers thus forms the fourth direction of basic research of the project.

4. Application Domains

4.1. Introduction

Constraint programming deals with the resolution of decision problems by means of rational, logical and computational techniques. Above all, constraint programming is founded on a clear distinction between, on the one hand the description of the constraints intervening in a problem, and on the other hand the techniques used for the resolution. The ability of constraint programming to handle in a flexible way heterogeneous constraints has raised the commercial interest for this paradigm in the early nighties. Among his fields of predilection, one finds traditional applications such as computer aided decision-making, scheduling, planning, placement, logistics or finance, as well as applications such as electronic circuits design (simulation, checking and test), DNA sequencing and phylogeny in biology, configuration of manufacturing products or web sites, formal verification of code.

4.2. Panorama

In 2015 the TASC team was involved in the following application domains:

- *Replanning* in industrial timetabling problems in a Labcom project with Eurodécision (see Figure 9).
- *Planning and replanning* in Data Centres taking into account energy consumption in the EPOC (Energy Proportional and Opportunistic Computing system) project.
- *Packing complex shapes* in the context of a warehouse (NetWMS2 project).
- Building decision support system for *resilient city development planning wrt climat change* (GRACeFUL project).
- *Optimizing electricity production* in the Gaspard Monge call program for Optimisation and Operation Research in the context of electricity production. In 2015 we were focussing on the systematic reformulation of time-series constraints for MIP solvers. This was done in order to integrate time-series constraints in existing integer linear programming models for electricity production.

5. Highlights of the Year

5.1. Highlights of the Year

5.1.1. Awards

Award at the MiniZinc Challenge 2016 solver competition in the Fixed category (Bronze). The aim of the challenge is to start to compare various constraint solving technology on the same problems sets. The focus is on finite domain propagation solvers. An auxiliary aim is to build up a library of interesting problem models, which can be used to compare solvers and solving technologies.

6. New Software and Platforms

6.1. AIUR

(Artificial Intelligence Using Randomness

FUNCTIONAL DESCRIPTION

The main idea is to be unpredictable by making some stochastic choices. The AI starts a game with a "mood" randomly picked up among 5 moods, dictating some behaviors (aggressive, fast expand, macro-game, ...). In addition, some other choices (productions, timing attacks, early aggressions, ...) are also taken under random conditions.

Learning is an essential part of AIUR . For this, it uses persistent I/O files system to record which moods are efficient against a given opponent, in order to modify the probability distribution for the mood selection. The current system allows both on-line and off-line learning.

- Contact: Florian Richoux
- URL: https://github.com/AIUR-group/AIUR

6.2. CHOCO

KEYWORDS: Constraint Programming - Scheduling - Optimisation - Operational research - Financial analysis - Planning

SCIENTIFIC DESCRIPTION

or 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, 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, a generic strategy based on explanations for large neighborhood search was designed and integrated within CHOCO. A corresponding paper appears at Constraints.

Within the context of the PhD thesis of Jean-Guillaume Fages, a documented package for graph variables was designed and integrated within CHOCO .

FUNCTIONAL DESCRIPTION

CHOCO is a Java discrete constraints library for describing hard combinatorial problems in the form of Constraint Satisfaction Problems and solving them with Constraint Programming techniques. Choco can be used to solve a broad range of real combinatorial problems. It is easy to use and offers excellent performance. This technique enables non-specialists to tackle strategic or operational problems, for instance, problems related to planning, scheduling, logistics, financial analysis and bio-informatics.

- Participants: Charles Prud'homme, Nicolas Beldiceanu, Jean-Guillaume Fages, Xavier Lorca, Thierry Petit and Rémi Douence
- Partner: Ecole des Mines de Nantes
- Contact: Julien Prud'homme
- URL: http://www.choco-solver.org/

6.3. GCCat

Global Constraint Catalog

KEYWORDS: Constraint Programming - Graph - Global constraint

FUNCTIONAL DESCRIPTION

The global constraint catalog presents and classifies global constraints and describes different aspects with meta data.

- Participants: Nicolas Beldiceanu and Sophie Demassey
- Contact: Nicolas Beldiceanu
- URL: http://sofdem.github.io/gccat/gccat/index.html

6.4. GCCat on time series

Global Constraint Catalog, Volume II, time-series constraints

KEYWORDS: Constraint Programming - Sequence - Transducer - Global constraint

FUNCTIONAL DESCRIPTION

The second volume of the Global Constraint Catalogue is devoted to time-series constraints. Within the context of Constraint Programming, time-series constraints go back to the work of Goldin and Kanellakis. This volume contains 626 constraints, which are explicitly described in terms of automata with accumulators. Checkers and propagators for all these constraints were synthesised from 22 transducers.

As in the first volume, the global constraints described in this second volume are not only accessible to humans, who can read the catalogue when searching for some information. It is also available to machines, which can read and interpret it. This is why there also exists an electronic version of this catalogue where one can get, for all time-series constraints, a complete description in terms of meta-data used in the first volume. In fact, unlike the first volume, *all the meta-data* of the electronic version as well as *all text and figures* of this second volume were automatically generated. While this second volume is by no means supposed to contain all possible time-series constraints, it contributes in the context of time-series constraints to the *systematic reconstruction* of the Global Constraint Catalogue that we have previously advocated. This reconstruction is based on the following methodology:

- First reuse, adapt or come up with abstractions, which allow to concisely represent structures and properties of time series as abstract combinatorial objects. In our context these abstractions essentially correspond to:
    1. Transducers where letters of the output alphabet are interpreted as semantic letters indicating how to recognise pattern occurrences.
    2. Transducers glue matrices expressing the relationship between the prefix, the suffix and the full sequence passed to a transducer.
    3. Properties associated to regular expressions corresponding to fragments of the input language of our transducers.
- Second, create from these abstract combinatorial objects a data base of concrete combinatorial objects.
- Third, synthesise concrete code for various technologies, languages, tasks from this data base of concrete combinatorial objects. In this context, correctness and efficiency of the synthesised code are essentially side product of:
    - The correctness of the formulae of our data base which is itself based on the wellformedness of our abstractions.
    - The generality behind our abstract combinatorial objects.

The time-series catalogue is done in the following way:

- All time-series constraints are now defined in a *compositional way* from a few basic constituents, i.e., patterns, features, aggregators, and predicates, which completely define the meaning of a constraint, where patterns are defined using regular expressions.
- Constraint names are now constructed in a systematic way as the *concatenation* of pattern name, feature name, and aggregation or predicate name.
- Given a pattern p, checkers and constraints are now *systematically synthesised* from a transducer that, given an input sequence over the input alphabet $\{<,=,>\}$, compares two adjacent values of a time-series and determines an output sequence over a output semantic alphabet describing how to recognise the occurrences of p.
- For each time-series constraint associated with a pattern p, the generation of an automaton with accumulators is completely driven by the transducer associated with pattern p as well as by *decoration tables* describing for each semantic letter of the output alphabet of the transducers how to generate accumulator updates. Code optimisation is ensured by using decoration tables that depend on properties of the pattern, of the feature, and of the aggregator associated with the time-series constraint.

- Lower and upper bounds of characteristics of time-series that appear in the restriction slot of a time-series constraint are synthesised from a *few parameterised formulae* that only depend on a restricted set of characteristics of the regular expression associated with the pattern.
- Parametrised glue matrices are provided for each transducer that corresponds to reversible time-series constraints. A concrete glue matrix is given for each reversible time-series constraint.
- Linear invariants are systematically obtained by applying the Farkas Lemma to the automata with accumulators that were synthesised. They consist of *linear constraints typically linking consecutive accumulator values*, e.g., see the legend of the second automaton of the constraints, which are generated even with non-linear accumulator updates. Missing linear invariants will be completed later on.
- Last but not least, time-series constraints were used for generating time-series verifying a conjunction of constraints both in the context of Constraint Programming and in the context of Linear Programming.
- In the context of sequential pattern mining, time-series constraint checkers can be used to identify and extract patterns from fixed sequences. While the time-series catalogue may need to be extended in order to capture more patterns, having a possibly large set of fixed time-series constraints is a natural safeguard to prevent overfitting when dealing with few sequences, at a price of not finding patterns that are not covered by the catalogue.
- Finally, both SICStus and MiniZinc code are synthesised. The later allows using time series constraints on many plate forms such as Choco, Gecode, ORtools, Cplex or Gurobi and is available Electronic Constraint Catalogue.
- Participants: Ekaterina Arafailova, Nicolas Beldiceanu, Rémi Douence, Mats Carlsson, Pierre Flener, Maria Andreina Francisco Rodriguez, Justin Pearson, Helmut Simonis
- Contact: Nicolas Beldiceanu
- URL: https://arxiv.org/abs/1609.08925

6.5. GHOST

General meta-Heuristic Optimization Solving Tool
FUNCTIONAL DESCRIPTION

GHOST is a template C++ library designed for StarCraft:BroodWartm. GHOST implements a meta-heuristic solver aiming to solve any kind of combinatorial and optimization RTS-related problems represented by a csp /cop. The solver handles dedicated geometric and assignment constraints in a way that is compatible with very strong real time requirements.

- Contact: Florian Richoux
- URL: http://github.com/richoux/GHOST

6.6. TorchCraft

Machine learning framework for games
FUNCTIONAL DESCRIPTION

TorchCraft is a library that enables deep learning research on Real-Time Strategy (RTS) games such as StarCraft: Brood War, by making it easier to control these games from a machine learning framework, here Torch. This white paper argues for using RTS games as a benchmark for AI research, and describes the design and components of TorchCraft.

- Participants: Gabriel Synnaeve, Nantas Nardelli, Alex Auvolat, Soumith Chintala, Timothe´e Lacroix, Zeming Lin, Florian Richoux, Nicolas Usunier
- Contact: Florian Richoux
- URL: https://arxiv.org/abs/1611.00625

7. New Results

7.1. Discrete Convexity

We introduce a propagator for pairs of Sum constraints, where the expressions in the sums respect a form of convexity. This propagator is parametric and can be instantiated for various concrete pairs, including Deviation, Spread, and the conjunction of Linear($\leq$) and Among. We show that despite its generality, our propagator (see Figure 1) is competitive in theory and practice with state-of-the-art propagators. (see AI journal paper).

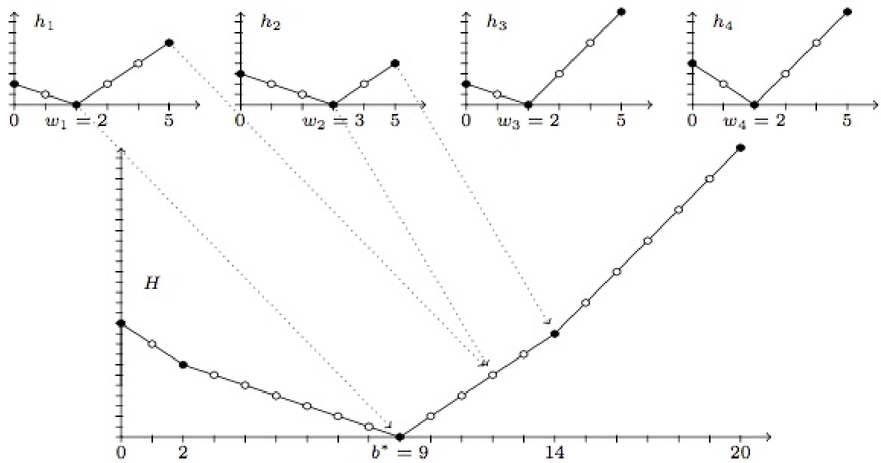


Figure 1. Illustration of the filtering wrt h function

7.2. Transducers

We describe a large family of constraints for structural time series by means of function composition. These constraints are on aggregations of features of patterns that occur in a time series, such as the number of its peaks, or the range of its steepest ascent. The patterns and features are usually linked to physical properties of the time series generator, which are important to capture in a constraint model of the system, i.e. a conjunction of constraints that produces similar time series. We formalise the patterns using finite transducers, whose output alphabet corresponds to semantic values that precisely describe the steps for identifying the occurrences of a pattern. Based on that description, we automatically synthesise automata with accumulators, as well as constraint checkers. The description scheme not only unifies the structure of the existing 30 time-series constraints in the Global Constraint Catalogue, but also leads to over 600 new constraints, with more than 100,000 lines of synthesised code. (see Constraint journal paper)

7.3. Compositional Glue Matrix and Bound for Time-Series Constraints

Integer time series are often subject to constraints on the aggregation of the integer features of all occurrences of some pattern within the series. For example, the number of inflexions may be constrained, or the sum of the peak maxima, or the minimum of the peak widths. It is currently unknown how to maintain domain consistency efficiently on such constraints. We propose parametric ways of systematically deriving glue constraints (see Figures 3 and 4 for the parametric and concrete glue constraints), which are a particular kind of implied

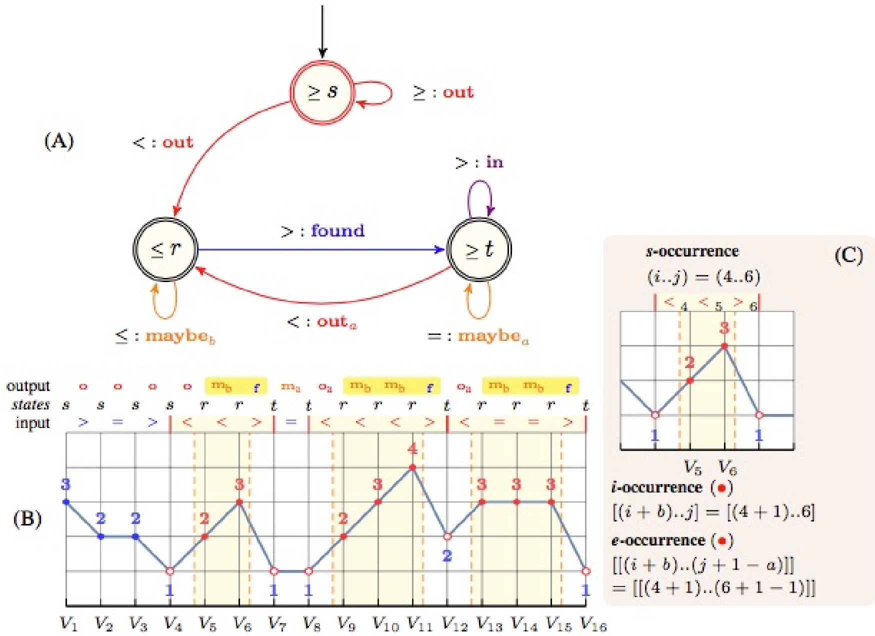


Figure 2. Transducer for the peak pattern and its execution on a sequence

constraints, as well as aggregation bounds (see Figure 5) that can be added to the decomposition of time-series constraints. We evaluate the beneficial propagation impact of the derived implied constraints and bounds, both alone and together. (see CP conference paper)

	s	r	t
s	$\phi_g(\overrightarrow{C}, \overleftarrow{C})$	$\phi_g(\overrightarrow{C}, \overleftarrow{C})$	$\phi_g(\overrightarrow{C}, \overleftarrow{C})$
r	$\phi_g(\overrightarrow{C}, \overleftarrow{C})$	$\phi_f(\overrightarrow{D}, \overleftarrow{D}, \delta_f^i)$	$\phi_f(\overleftarrow{C}, \overrightarrow{D}, \overleftarrow{D}, \delta_f^i)$
t	$\phi_g(\overrightarrow{C}, \overleftarrow{C})$	$\phi_f(\overleftarrow{C}, \overrightarrow{D}, \overleftarrow{D}, \delta_f^i)$	$\phi_g(\overrightarrow{C}, \overleftarrow{C})$

Figure 3. Parametrised glue matrix for the peak pattern expressed in term of parametrised functions depending on the states pairs between the prefix and the suffix of a sequence

7.4. Reformulation of time-series constraint in MIP

A checker for a constraint on a variable sequence can often be compactly specified by an automaton, possibly with accumulators, that consumes the sequence of values taken by the variables; such an automaton can also be used to decompose its specified constraint into a conjunction of logical constraints. The inference achieved by this decomposition in a CP solver can be boosted by automatically generated implied constraints on the accumulators, provided the latter are updated in the automaton transitions by linear expressions. Automata with non-linear accumulator updates can be automatically synthesised for a large family of time-series constraints.

	s	r	t
s	$\overrightarrow{C}+\overleftarrow{C}$	$\overrightarrow{C}+\overleftarrow{C}$	$\overrightarrow{C}+\overleftarrow{C}$
r	$\overrightarrow{C}+\overleftarrow{C}$	1	1
t	$\overrightarrow{C}+\overleftarrow{C}$	1	$\overrightarrow{C}+\overleftarrow{C}$

Figure 4. Concrete glue matrix for the number peak constraint expressed in term of concrete functions depending on the states pairs between the prefix and the suffix of a sequence

$\mathtt{sv} \leq 3 \vee \mathtt{rv} \leq 1 \Rightarrow \mathtt{VALUE} = 0$
$\mathtt{VALUE} \geq 0$
$\mathtt{rv} = 2 \Rightarrow \mathtt{VALUE} \leq \lfloor \mathtt{sv}/4 \rfloor$ ①
$\mathtt{rv} \geq 3 \Rightarrow \mathtt{VALUE} \leq \mathtt{max}(0, \lfloor (\mathtt{sv}-1)/3 \rfloor)$ ②
$\mathtt{required}(\mathtt{VARIABLES}, \mathtt{var})$
where
$\mathtt{sv} = |\mathtt{VARIABLES}|$
$\mathtt{rv} = \mathtt{range}(\mathtt{VARIABLES.var})$

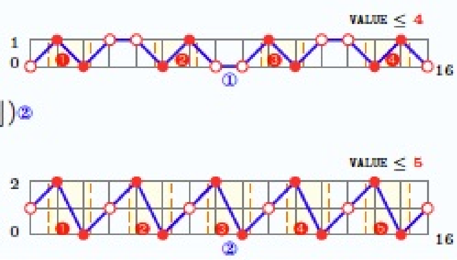


Figure 5. Upper bound on the number of zigzag depending on the domain range being equal to 2 or greater than or equal to 3

In this paper, we describe and evaluate extensions to those techniques. First, we improve the automaton synthesis to generate automata with fewer accumulators. Second, we decompose a constraint specified by an automaton with accumulators into a conjunction of linear inequalities, for use by a MIP solver. Third, we generalise the implied constraint generation to cover the entire family of time-series constraints. The newly synthesised automata for time-series constraints outperform the old ones, for both the CP and MIP decompositions, and the generated implied constraints boost the inference, again for both the CP and MIP decompositions. We evaluate CP and MIP solvers on a prototypical application modelled using time-series constraints. (see CPAIOR conference paper)

7.5. Scheduling Constraint for Video Summarisation

Given a sequence of tasks T subject to precedence constraints between adjacent tasks, and given a set of fixed intervals I, the TaskIntersection (T,I,o,inter) constraint restricts the overall intersection of the tasks of T with the fixed intervals of I to be greater than or equal or less than or equal to a given limit inter. We provide a bound(Z)-consistent cost filtering algorithm wrt the starts and the ends of the tasks for the TaskIntersection constraint and evaluate the constraint on the video summarisation problem. (see CPAIOR conference paper)

7.6. A Model Seeker for Learning Constraints Models from Positive Samples

We describe a system which generates finite domain constraint models from positive example solutions (e.g. see Figure 6 giving a season schedule of the Bundesliga), for highly structured problems. The system is based on the global constraint catalog, providing the library of constraints that can be used in modeling, and the Constraint Seeker tool, which finds a ranked list of matching constraints given one or more sample call patterns (e.g. see Figure 7 giving the model learned for the input data of Figure 6). We have tested the modeler with 230 examples, ranging from 4 to 6,500 variables, using between 1 and 7,000 samples. These examples come from a variety of domains, including puzzles, sports-scheduling, packing and placement, and design theory. When comparing against manually specified canonical models for the examples, we achieve a hit rate of 50 percent, processing the complete benchmark set in less than one hour on a laptop. Surprisingly, in many cases the system finds usable candidate lists even when working with a single, positive example. (see Book chapter of Data Mining and Constraint Programming)

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
8	1	14	11	4	7	2	15	12	13	6	9	10	3	18	5	16	17
3	14	17	2	13	6	5	12	9	16	11	18	1	4	15	8	7	10
...																	
18	17	2	1	4	3	6	5	10	9	16	15	14	13	12	11	8	7
13	12	11	14	17	16	15	2	9	6	1	8	7	4	5	18	3	10
...																	

Figure 6. Input data corresponding to a flat sample (a sequence of integer values) giving a one year season schedule of the Bundesliga

7.7. Global Constraint Catalog Volume II: Time-Series Constraints

First this report presents a restricted set of 22 finite transducers used to synthesise structural time-series constraints described by means of a multi-layered function composition scheme. Second it provides the

-	Sequence Generator	Projection	Constraint Conjunction
1	scheme(612,34,18,34,1)	id	alldifferent*18
2	scheme(612,34,18,2,2)	id	alldifferent*153
3	scheme(612,34,18,1,18)	id	alldifferent*34
4	scheme(612,34,18,1,18)	absolute_value	symmetric_alldifferent([1..18])*34
5	scheme(612,34,18,17,1)	absolute_value	alldifferent*36
6	repart(612,34,18,34,9)	id	sum_ctr(0)*306
7	repart(612,34,18,34,9)	id	twin*1
8	repart(612,34,18,34,9)	id	elements([i,-i])*1
9	first(9,[1,3,5,7,9,11,13,15,17])	id	strictly_increasing*1
10	vector(612)	id	global_cardinality([-18.. -1-17,0-0,1..18-17])*1
11	repart(612,34,18,34,9)	id	sum_powers5_ctr(0)*306
12	repart(612,34,18,34,9)	id	sum_cubes_ctr(0)*306
13	repart(612,34,18,34,3)	sign	global_cardinality([-1-3,0-0,1-3])*102
14	scheme(612,34,18,34,1)	sign	global_cardinality([-1-17,0-0,1-17])*18
15	repart(612,34,18,17,9)	sign	global_cardinality([-1-2,0-0,1-2])*153
16	repart(612,34,18,2,9)	sign	global_cardinality([-1-17,0-0,1-17])*18
17	scheme(612,34,18,1,18)	sign	global_cardinality([-1-9,0-0,1-9])*34
18	repart(612,34,18,34,9)	sign	sum_ctr(0)*306
19	repart(612,34,18,34,9)	sign	twin*1
20	repart(612,34,18,34,9)	absolute_value	twin*1
21	repart(612,34,18,34,9)	sign	elements([i,-i])*1
22	scheme(612,34,18,34,1)	sign	among_seq(3,[-1])*18
23	repart(612,34,18,34,9)	absolute_value	elements([i,i])*1
24	first(9,[1,3,5,7,9,11,13,15,17])	absolute_value	strictly_increasing*1
25	first(6,[1,4,7,10,13,16])	absolute_value	strictly_increasing*1
26	scheme(612,34,18,34,1)	absolute_value	nvalue(17)*18

Figure 7. Model, i.e. conjunction of global constraints, learned from the single flat sample

corresponding synthesised catalogue of structural time-series constraints where each of the 626 constraints is explicitly described in terms of automata with accumulators, see Figure 8 for the syntesised automaton of the sum surf peak constraint. (arXiv 1609.08925)

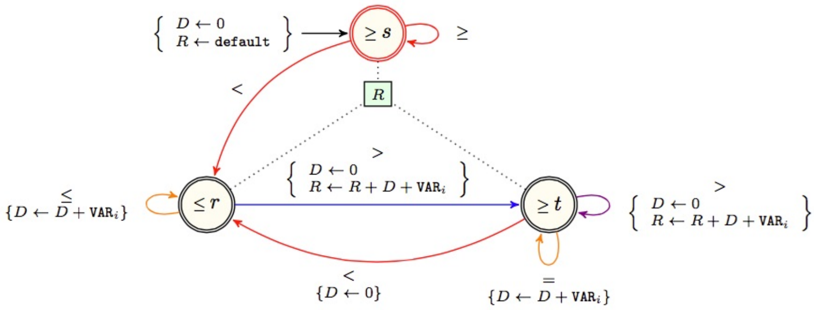


Figure 8. Synthesised automaton with accumulator of the sum surf peak constraint obtained from the transducer of the peak pattern

7.8. Probabilistic Model for Binary CSP

This work introduces a probabilistic-based model for binary CSP that provides a fine grained analysis of its internal structure. Assuming that a domain modification could occur in the CSP, it shows how to express, in a predictive way, the probability that a domain value becomes inconsistent, then it express the expectation of the number of arc-inconsistent values in each domain of the constraint network. Thus, it express the expectation of the number of arc-inconsistent values for the whole constraint network. Next, it provides bounds for each of these three probabilistic indicators. Finally, a polytime algorithm, which propagates the probabilistic information, is presented. (see arXiv 1606.03894 or [19])

7.9. Estimating parallel runtimes for randomized algorithms in constraint solving

We present a detailed analysis of the scalability and parallelisation of Local Search algorithms for constraint-based and SAT (Boolean satisfiability) solvers. We propose a framework to estimate the parallel performance of a given algorithm by analyzing the runtime behavior of its sequential version. Indeed, by approximating the runtime distribution of the sequential process with statistical methods, the runtime behavior of the parallel process can be predicted by a model based on order statistics. We apply this approach to study the parallel performance of a constraint-based Local Search solver (Adaptive Search), two SAT Local Search solvers (namely Sparrow and CCASAT), and a propagation-based constraint solver (Gecode, with a random labeling heuristic). We compare the performance predicted by our model to actual parallel implementations of those methods using up to 384 processes. We show that the model is accurate and predicts performance close to the empirical data. Moreover, as we study different types of problems, we observe that the experimented solvers exhibit different behaviors and that their runtime distributions can be approximated by two types of distributions: exponential (shifted and non-shifted) and lognormal. Our results show that the proposed framework estimates the runtime of the parallel algorithm with an average discrepancy of 21 percent w.r.t. the empirical data across all the experiments with the maximum allowed number of processors for each technique. (see Journal of Heuristics)

7.10. ghost: A Combinatorial Optimization Framework for Real-Time Problems

We presents GHOST, a combinatorial optimization framework that a real-time strategy (RTS) AI developer can use to model and solve any problem encoded as a constraint satisfaction/optimization problem (CSP/COP). We show a way to model three different problems as a CSP/COP, using instances from the RTS game StarCraft as test beds. Each problem belongs to a specific level of abstraction (the target selection as reactive control problem, the wall-in as a tactics problem, and the build order planning as a strategy problem). In our experiments, GHOST shows good results computed within some tens of milliseconds. We also show that GHOST outperforms state-of-the-art constraint solvers, matching them on the resources allocation problem, a common combinatorial optimization problem. (see IEEE Transactions on Computational Intelligence and AI in games journal)

7.11. TorchCraft: a Library for Machine Learning Research on Real-Time Strategy Games

We present TorchCraft, a library that enables deep learning research on Real-Time Strategy (RTS) games such as StarCraft: Brood War, by making it easier to control these games from a machine learning framework, here Torch. This white paper argues for using RTS games as a benchmark for AI research, and describes the design and components of TorchCraft. (see arXiv 1611.00625)

7.12. POSL: A Parallel-Oriented metaheuristic-based Solver Language

For a couple of years, all processors in modern machines are multi-core. Massively parallel architectures, so far reserved for super-computers, become now available to a broad public through hardware like the Xeon Phi or GPU cards. This architecture strategy has been commonly adopted by processor manufacturers, al- lowing them to stick with Moore's law. However, this new architecture implies new ways to design and implement algorithms to exploit its full potential. This is in particular true for constraint-based solvers dealing with combinatorial optimiza- tion problems. Here we propose a Parallel-Oriented Solver Language (POSL, pronounced "puzzle"), a new framework to build interconnected meta-heuristic based solvers working in parallel. The novelty of this approach lies in looking at solver as a set of components with specific goals, written in a parallel-oriented language based on operators. A major feature in POSL is the possibility to share not only information, but also behaviors, allowing solver modifications during runtime. Our framework has been designed to easily build constraint-based solvers and reduce the developing effort in the context of parallel architecture. POSL's main advan- tage is to allow solver designers to quickly test different heuristics and parallel com- munication strategies to solve combinatorial optimization problems, usually time- consuming and very complex technically, requiring a lot of engineering.

7.13. Towards Automated Strategies in Satisfiability Modulo Theory

SMT solvers include many heuristic components in order to ease the theorem proving process for different logics and problems. Handling these heuristics is a non-trivial task requiring specific knowledge of many theories that even a SMT solver developer may be unaware of. This is the first barrier to break in order to allow end-users to control heuristics aspects of any SMT solver and to successfully build a strategy for their own purposes. We present a first attempt for generating an automatic selection of heuristics in order to improve SMT solver efficiency and to allow end-users to take better advantage of solvers when unknown problems are faced. Evidence of improvement is shown and the basis for future works with evolutionary and/or learning-based algorithms are raised (see Genetic Programming conference paper).

7.14. Using CP for the Urban Transit Crew Rescheduling Problem

Scheduling urban and trans-urban transportation is an important issue for industrial societies. The Urban Transit Crew Scheduling Problem is one of the most important optimization problem related to this issue. It mainly relies on scheduling bus drivers workday respecting both collective agreements (see Figure 9 for an example of regulation rule) and the bus schedule needs. If this problem has been intensively studied from a tactical point of view, its operational aspect has been neglected while the problem becomes more and more complex and more and more prone to disruptions. In this way, this paper presents how the constraint programming technologies are able to recover the tactical plans at the operational level in order to efficiently help in answering regulation needs after disruptions (see CP conference paper).

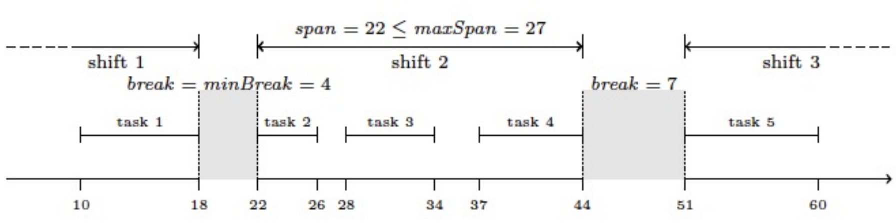


Figure 9. Illustration of typical regulation rule in the Labcom project; Shift 2 of an employee is composed of tasks 2, 3, and 4. It is between shifts 1 and 3 of the same employee. A break of duration 4 is scheduled between task 1 of shift 1 and task 2 of shift 2, because the gap between these tasks is at least a break of 4. Similarly, a break of duration 7 is scheduled after task 4 of shift 2 and task 5 of shift 3. No other breaks can be scheduled between the tasks of shift 2 because of the minimum break duration. The span of shift 2 is 22 and does not exceed 27: it is composed of tasks 2, 3, and 4, as well as of the two gaps between these tasks.

7.15. Traveling salesman and the tree: the importance of search in CP

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 paper 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. (see Constraints journal)

7.16. Event Selection Rules to Compute Explanations

Explanations have been introduced in the previous century. Their interest in reducing the search space is no longer questioned. Yet, their efficient implementation into CSP solver is still a challenge. In this paper, we introduce ESeR, an Event Selection Rules algorithm that filters events generated during propagation. This dynamic selection enables an efficient computation of explanations for intelligent backtracking algorithms. We show the effectiveness of our approach on the instances of the last three MiniZinc challenges. (see arXiv 1608.08015 or [20])

7.17. Towards energy-proportional Clouds partially powered by renewable energy

With the emergence of the Future Internet and the dawning of new IT models such as cloud computing, the usage of data centers (DC), and consequently their power consumption, increase dramatically. Besides the ecological impact, the energy consumption is a predominant criterion for DC providers since it determines the daily cost of their infrastructure. As a consequence, power management becomes one of the main challenges for DC infrastructures and more generally for large-scale-distributed systems. In this paper, we present the EpoCloud prototype, from hardware to middleware layers. This prototype aims at optimizing the energy consumption of mono-site Cloud DCs connected to the regular electrical grid and to renewable-energy sources (see Journal of Computing).

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Contracts with Industry

8.1.1. Labcom TransOp

Participants: Charles Prud'Homme, Xavier Lorca.

Title: TransOp.

Duration: 2014-2016.

Type: **ongoing project**.

Others partners: Eurodécision.

The goal of the project is to handle robustness in the context of industrial timetabling problems with constraint programming using CHOCO. The project is managed by Xavier Lorca.

8.2. Bilateral Grants with Industry

8.2.1. Gaspard Monge

Participants: Nicolas Beldiceanu, Helmut Simonis.

Title: Gaspard Monge 3.

Duration: 2016.

Type: **ongoing project**.

Others partners: EDF.

Within the context of the Gaspard Monge call program for Optimisation and Operation Research, we work with EDF on the research initiative on *Optimization and Energy*. The goal of the project (continuation of last years projects) is to provide a systematic reformulation of time-series constraints in term of linear constraints that can be used in a MIP solver.

9. Partnerships and Cooperations

9.1. Regional Initiatives

9.1.1. EPOC

With the emergence of the Future Internet and the dawning of new IT models such as cloud computing, the usage of data centers (DC), and consequently their power consumption, increase dramatically. Besides the ecological impact, the energy consumption is a predominant criteria for DC providers since it determines the daily cost of their infrastructure. As a consequence, power management becomes one of the main challenges for DC infrastructures and more generally for large-scale distributed systems. In this paper, the EPOC project which focuses on optimising the energy consumption of mono-site DCs connected to the regular electrical grid and to renewable energy sources.

9.1.2. SmartCat

Participants: Eric Monfroy, Charlotte Truchet.

Title: Online optimization for chemical reactions.

Others partners: CEISAM.

The SmartCat project, started in 2015 on regional fundings, aims at developing an intelligent automatised tool for online chemistry. Contrarily to the traditional batch chemistry, where reactants are mixed in a glass, online chemistry consists in having a flow of reactants in a tube, possibly passing through ovens are pressure control mechanisms. This way, the reaction happens continuously and it can produce much more products within a system of reasonable size. SmartCat integrates a controller for which intelligent tools need to be developed. These tools will analyse the product of the reaction and adapt the conditions (stoechiometry, pressure, temperature, catalysis) in order to optimise the yield. TASC contributes to this project by developing these methods, based on local search techniques.

9.1.3. Atlanstic 1

Participant: Florian Richoux.

Title: Atlanstic project about deep learning for games.

Duration: 2016.

Topic: deep learning for games.

9.1.4. Atlanstic 2

Participant: Charles Prud Homme.

Title: CoMe4ACloud.

Duration: 2016.

Topic: CoMe4ACloud is an Atlanstic2020 funded project whose objective is to provide an end-to-end solution for autonomic Cloud services. To that end, we rely on techniques of Constraint Programming so as a decision-making tool and Model-driven Engineering to ease the automatic generation of the so-called autonomic managers as well as their communication with the managed system (see Constraints and Model Engineering for Autonomic Clouds). The project is led by ASCOLA research team and involves also AtlanModels and TASC.

9.2. National Initiatives

9.2.1. ANR NetWMS2

Participants: Gilles Chabert, Ignacio Salas Donoso, Nicolas Beldiceanu.

Title: Networked Warehouse Management Systems 2: packing with complex shapes.

Duration: 2011-2014.

Type: cosinus research program.

Budget: 189909 Euros.

Others partners: KLS Optim and CONTRAINTES (Inria Rocquencourt).

This project builds on the former European FP6 Net-WMS Strep project that has shown that constraint-based optimisation techniques can considerably improve industrial practice for box packing problems, while identifying hard instances that cannot be solved optimally, especially in industrial 3D packing problems with rotations, the needs for dealing with more complex shapes (e.g. wheels, silencers) involving continuous values. This project aims at generalizing the geometric kernel *geost* for handling non-overlapping constraints for complex two and three dimensional curved shapes as well as domain specific heuristics. This will be done within the continuous solver IBEX, where discrete variables will be added for handling polymorphism (i.e., the fact that an object can take one shape out of a finite set of given shapes). A filtering algorithm has been devised in the case of objects described by nonlinear inequalities and is now under testing with the Ibex library. This work has been presented in a workshop on interval methods & geometry in ENSTA Bretagne.

9.3. European Initiatives

9.3.1. FP7 & H2020 Projects

Within the context of the First Future and Emerging Technologies (FET) Proactive projects under Horizon 2020 Framework Programme the GRACeFUL project started this year. From an application point of view the project develops scalable rapid assessment tools for collective policy making in global systems, and test these on climate-resilient urban design. From a technical point of view it provides domain specific languages that are embedded in functional programming and constraint programming languages. Within the project TASC is responsible for the constraint part. To interact with policy makers it uses some qualitative network model (see Figure 10) embedded with constraint programming models that also capture dependancy between potential actions as well as costs.

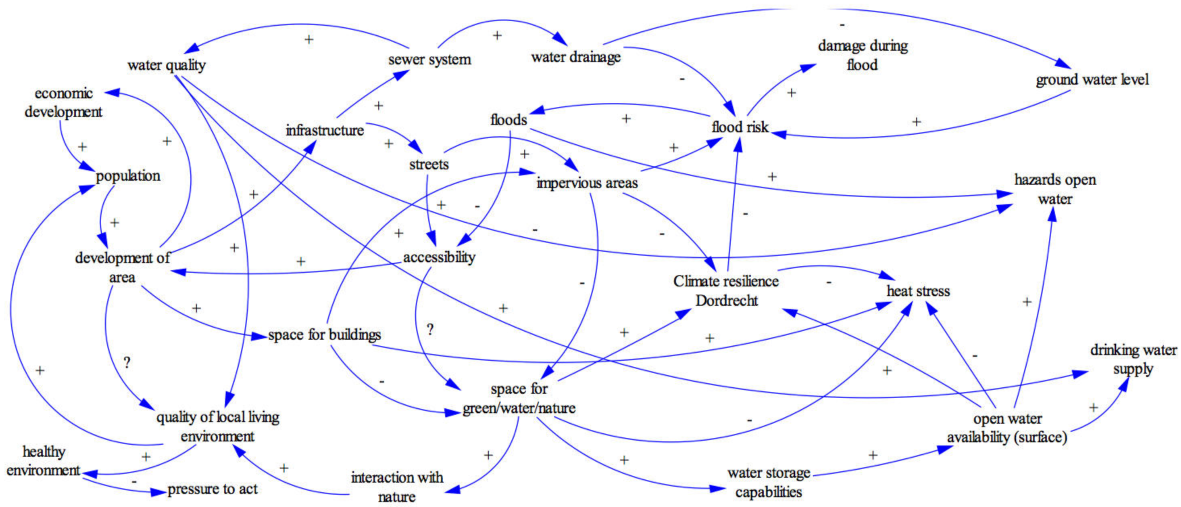


Figure 10. Illustration of some qualitative network capturing causality in the context of flooding prevention

9.4. International Initiatives

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

9.4.1.1. TASCMELB

Title: Synergy between Filtering and Explanations for Scheduling and Placement Constraints

International Partner (Institution - Laboratory - Researcher):

> NICTA (Australia) - Optimisation Research Group (Optimisation) - Pascal van Hentenryck

Start year: 2014

See also: http://www.normalesup.org/~truchet/TASCMELB.html

In the context of Constraint Programming and SAT the project addresses the synergy between filtering (removing values from variables) and explanations (explaining why values were removed in term of clauses) in order to handle in a more efficient way correlated resource scheduling and placement constraints. It combines the strong point of Constraint Programming, namely removing value that leads to infeasibility, with the strong point of SAT, namely taking advantage from past failure in order to quickly identify infeasible sub-problems. In 2016 we got the following the following new result *using rewriting for synthesising filtering algorithm for the Allen constraint*: For all 8192 combinations of Allen's 13 relations between one task with origin oi and fixed length li and another task with origin oj and fixed length lj, we give a formula evaluating to a set of integers which are infeasible for a task origin for the given combination. Such forbidden regions are useful e.g. in a range-consistency maintaining propagator for an Allen constraint in finite domain constraint programming. No visit to Melbourne was done this year because of VISA problem. Consequently we also did remotly (i.e. from Nantes) the following result: the availability of the time-series constraints of the time-series constraint catalog available in the MiniZinc modelling language (and consequently made them accessible to solvers like Choco or Cplex).

9.5. International Research Visitors

9.5.1. Visits of International Scientists

- A visit regarding time-series constraints of Andreina Francisco RodriguezHelmut Simonis, Pierre Flener and Justin Pearson in Nantes in May.
- A visit regarding time-series constraints of Helmut Simonis, in July in May.

9.5.2. Visits to International Teams

- Two visits of E. Arafailova regarding time-series constraints in Cork (March 2016) and in Uppsala (April 2016)
- Three visits of N. Beldiceanu regarding time-series constraint in Cork (June 2016) and in Uppsala (February 2016, August 2016)

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Leadership within the Scientific Community

Charlotte Truchet was elected in the ACP committee.

10.1.2. Research Administration

Preparation by the whole team of the two evaluations that respectively toke place in January 2016 (hceres evaluation) and in March 2016 (inria evaluation).

10.2. Teaching - Supervision - Juries

10.2.1. Supervision

PhD: Ignacio Salas Donoso, Packing curved objects with interval methods, Started in May 2013, PhD defense April 5 2016 with committee Luc Jaulin, Gilles Trombettoni, François Fages, see PhD thesis, Gilles Chabert and Nicolas Beldiceanu.

PhD in progress : Gilles Madi Wamba, Mixing constraint programming and behavioural models to manage energy consumption in data centre, October 2014, Nicolas Beldiceanu and Didier Lime.

PhD in progress : Alejandro Reyes Amaro, Toward autonomous parallel algorithms for constraint-based problems, October 2014, Eric Monfroy and Florian Richoux.

PhD in progress : Anicet Bart, Solving mixed constraints, application to the management of mobile sensors, October 2014, Eric Monfroy and Charlotte Truchet.

PhD in progress : Ekaterina Arafailova, Functional constraints, September 2015, Nicolas Beldiceanu and Rémi Douence.

PhD in progress : Nicolas Galvez, Hybrid Algorithms for Search Based Software Engineering, December 2014, Eric Monfroy with Frédéric Saubion from Angers University and C. Castro from UTFSM Valparaiso, Chili.

10.3. Popularization

- Maintenance of the global constraint catalogue.
- Illustrations of the volume II of the global constraint catalogue: 2000 figures.

11. Bibliography

Publications of the year

Articles in International Peer-Reviewed Journal

[1] N. Beldiceanu, M. Carlsson, R. Douence, H. Simonis. *Using finite transducers for describing and synthesising structural time-series constraints*, in "Constraints", January 2016, vol. 21, n^{o} 1 [*DOI :* 10.1007/s10601-015-9200-3], https://hal.inria.fr/hal-01370322.

[2] N. Beldiceanu, B. Dumas Feris, P. Gravey, S. Hasan, C. Jard, T. Ledoux, Y. Li, D. Lime, G. Madi-Wamba, J.-M. Menaud, P. Morel, M. Morvan, M.-L. MOULINARD, A.-C. Orgerie, J.-L. Pazat, O. H. Roux, A. Sharaiha. *Towards energy-proportional Clouds partially powered by renewable energy*, in "Computing", January 2017, vol. 99, n^{o} 1, 20 [*DOI :* 10.1007/s00607-016-0503-z], https://hal.inria.fr/hal-01340318.

[3] J.-G. Fages, X. Lorca, L.-M. Rousseau. *The salesman and the tree: the importance of search in CP*, in "Constraints", 2016, vol. 21, p. 145 - 162 [*DOI :* 10.1007/s10601-014-9178-2], https://hal.archives-ouvertes.fr/hal-01436214.

[4] J.-N. Monette, N. Beldiceanu, P. Flener, J. Pearson. *A Parametric Propagator for Pairs of Sum Constraints with a Discrete Convexity Property*, in "Artificial Intelligence", 2016, vol. 241, p. 170 - 190 [*DOI :* 10.1016/j.artint.2016.08.006], https://hal.inria.fr/hal-01374714.

[5] F. Richoux, A. Uriarte, J.-F. Baffier. *GHOST: A Combinatorial Optimization Framework for Real-Time Problems*, in "IEEE Transactions on Computational Intelligence and AI in games", 2016, vol. 8, n^{o} 4, p. 377-388 [*DOI :* 10.1109/TCIAIG.2016.2573199], https://hal.archives-ouvertes.fr/hal-01436124.

International Conferences with Proceedings

[6] E. Arafailova, N. Beldiceanu, R. Douence, P. Flener, M. A. Francisco Rodríguez, J. Pearson, H. Simonis. *Time-Series Constraints: Improvements and Application in CP and MIP Contexts*, in "CPAIOR 2016 - 13th International Conference on Integration of Artificial Intelligence and Operations Research Techniques in Constraint Programming", Banff, Canada, C.-G. Quimper (editor), Lecture Notes in Computer Science, Springer, May 2016, vol. 9676, p. 18-34 [*DOI :* 10.1007/978-3-319-33954-2], https://hal.inria.fr/hal-01355262.

[7] X. Lorca, C. Prud 'homme, A. Questel, B. Rottembourg.*Using Constraint Programming for the Urban Transit Crew Rescheduling Problem*, in "Principles and Practice of Constraint Programming", Toulouse, France, September 2016, vol. 9892, p. 636 - 649 [*DOI :* 10.1007/978-3-319-44953-1_40], https://hal.archives-ouvertes.fr/hal-01436288.

National Conferences with Proceeding

[8] A. Reyes Amaro, E. Monfroy, F. Richoux.*Étude de stratégies parallèles de coopération avec POSL*, in "Douzièmes Journées Francophones de Programmation par Contraintes (JFPC)", Montpellier, France, June 2016, https://hal.archives-ouvertes.fr/hal-01436130.

Conferences without Proceedings

[9] A. Bart, C. Truchet, E. Monfroy.*Contraintes sur des flux appliquées a la vérification de programmes audio*, in "Onzièmes Journées Francophones de Programmation par Contraintes", Bordeaux, France, June 2016, https://hal.archives-ouvertes.fr/hal-01234179.

[10] G. Madi Wamba, N. Beldiceanu.*The TaskIntersection Constraint*, in "Integration of AI and OR techniques in constraint programming", Banff, Canada, Integration of AI and OR techniques in constraint programming, Springer, May 2016, vol. 9676, p. 246-261 [*DOI :* 10.1007/978-3-319-33954-2_18], https://hal.archives-ouvertes.fr/hal-01436044.

Scientific Books (or Scientific Book chapters)

[11] N. Beldiceanu, H. Simonis.*ModelSeeker: Extracting Global Constraint Models from Positive Examples*, in "Data Mining and Constraint Programming", Data Mining and Constraint Programming, Springer, December 2016, vol. 10101, p. 77-95 [*DOI :* 10.1007/978-3-319-50137-6_4], https://hal.archives-ouvertes.fr/hal-01436062.

[12] D. Churchill, M. Preuss, F. Richoux, G. Synnaeve, A. Uriarte, S. Ontañón, M. Certicky. *StarCraft Bots and Competitions*, in "Encyclopedia of Computer Graphics and Games", N. Lee (editor), Springer International Publishing, 2016, p. 1-18 [*DOI :* 10.1007/978-3-319-08234-9_18-1], https://hal.archives-ouvertes.fr/hal-01436110.

[13] A. Reyes Amaro, E. Monfroy, F. Richoux. *POSL: A Parallel-Oriented metaheuristic-based Solver Language*, in "Springer Recent Developments of Metaheuristics", Springer, 2016, https://hal.archives-ouvertes.fr/hal-01436119.

Scientific Popularization

[14] A. Aggoun, N. Beldiceanu, G. Chabert, F. Fages.*Packing with Complex Shapes*, in "ERCIM News", 2016, n° 105, 32, https://hal.archives-ouvertes.fr/hal-01378471.

Other Publications

[15] E. Arafailova, N. Beldiceanu, M. Carlsson, P. Flener, M. A. Francisco Rodríguez, J. Pearson, H. Simonis.*Systematic Derivation of Bounds and Glue Constraints for Time-Series Constraints*, August 2016, slides corresponding to this paper [*DOI :* 10.1007/978-3-319-44953-1_2], https://hal.inria.fr/hal-01370317.

[16] R. DOUENCE, P. FLENER, M. A. F. RODRÍGUEZ, J. PEARSON, H. SIMONIS, M. A. FRANCISCO RODRÍGUEZ, M. CARLSSON, N. BELDICEANU.*Global Constraint Catalog, Volume II, Time-Series Constraints*, September 2016, 2709, Synthesized catalogue (text, figures, code) of time-series constraints, https://hal.inria.fr/hal-01374721.

[17] A. REYES-AMARO, E. MONFROY, F. RICHOUX.*Suplementary document for Studying parallel communication strategies with POSL*, July 2016, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01345775.

[18] G. SYNNAEVE, N. NANTAS, A. ALEX, S. CHINTALA, L. TIMOTHÉE, L. ZEMING, F. RICHOUX, N. USUNIER.*TorchCraft: a Library for Machine Learning Research on Real-Time Strategy Games*, January 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01436134.

References in notes

[19] A. BALAFREJ, X. LORCA, C. TRUCHET.*A Probabilistic-Based Model for Binary CSP*, in "CoRR", 2016, vol. abs/1606.03894, http://arxiv.org/abs/1606.03894.

[20] C. PRUD'HOMME, X. LORCA, N. JUSSIEN.*Event Selection Rules to Compute Explanations*, in "CoRR", 2016, vol. abs/1608.08015, http://arxiv.org/abs/1608.08015.

Project-Team TEA

Time, Events and Architectures

RESEARCH CENTER
Rennes - Bretagne-Atlantique

THEME
Embedded and Real-time Systems

Table of contents

Project-Team TEA

Creation of the Team: 2014 January 01, updated into Project-Team: 2015 January 01

Keywords:

Computer Science and Digital Science:

1.2. - Networks
1.2.7. - Cyber-physical systems
1.2.8. - Network security
1.5. - Complex systems
1.5.1. - Systems of systems
1.5.2. - Communicating systems
2.1. - Programming Languages
2.1.1. - Semantics of programming languages
2.1.6. - Concurrent programming
2.1.8. - Synchronous languages
2.1.10. - Domain-specific languages
2.2. - Compilation
2.2.1. - Static analysis
2.2.3. - Run-time systems
2.3. - Embedded and cyber-physical systems
2.3.1. - Embedded systems
2.3.2. - Cyber-physical systems
2.3.3. - Real-time systems
2.4. - Verification, reliability, certification
2.4.1. - Analysis
2.4.2. - Model-checking
2.4.3. - Proofs
2.5. - Software engineering
4.4. - Security of equipment and software
4.5. - Formal methods for security
4.7. - Access control
5.7.2. - Music
6.1.1. - Continuous Modeling (PDE, ODE)
6.1.3. - Discrete Modeling (multi-agent, people centered)
6.2.1. - Numerical analysis of PDE and ODE
6.2.5. - Numerical Linear Algebra
6.2.6. - Optimization
7.4. - Logic in Computer Science
7.6. - Computer Algebra

Other Research Topics and Application Domains:

5.1. - Factory of the future
5.2. - Design and manufacturing

6.1.1. - Software engineering
6.4. - Internet of things
6.6. - Embedded systems
9.2.1. - Music, sound
9.4.1. - Computer science

1. Members

Research Scientists

Jean-Pierre Talpin [Inria, Team leader, Senior Researcher, HDR]
Thierry Gautier [Inria, Senior Researcher]
Vania Joloboff [Inria, Senior Researcher]

Technical Staff

Loïc Besnard [CNRS SED, Senior Research Engineer, seconded at 60%]
Clément Guy [Inria, until Aug 2016]
Alexandre Honorat [Inria]
Hai Nam Tran [Inria, from Dec 2016]

PhD Student

Simon Lunel [Mitsubishi Electrics R&D, granted by CIFRE]

Visiting Scientists

Rajesh Gupta [UC San Diego, Jul 2016]
Brian Larson [FDA/KSU, Jul 2016]

Administrative Assistants

Stéphanie Lemaile [Inria]
Armelle Mozziconacci [CNRS, from May 2016]

Other

Daian Yue [Inria, Master Intern, from Feb 2016 until Jun 2016]

2. Overall Objectives

2.1. Introduction

An embedded architecture is an artifact of heterogeneous constituents and at the crossing of several design viewpoints: software, embedded in hardware, interfaced with the physical world. Time takes different forms when observed from each of these viewpoints: continuous or discrete, event-based or time-triggered. Unfortunately, modeling and programming formalisms that represent software, hardware and physics significantly alter this perception of time. Moreover, time reasoning in system design is usually isolated to a specific design problem: simulation, profiling, performance, scheduling, parallelization, simulation. The aim of project-team TEA is to define conceptually unified frameworks for reasoning on composition and integration in cyber-physical system design, and to put this reasoning to practice by revisiting analysis and synthesis issues in real-time system design with soundness and compositionality gained from formalization.

2.2. Context

In the construction of complex systems, information technology (IT) has become a central force of revolutionary changes, driven by the exponential increase of computational power. In the field of telecommunication, IT provides the necessary basis for systems of networked distributed applications. In the field of control engineering, IT provides the necessary basis for embedded control applications. The combination of telecommunication and embedded systems into networked embedded systems opens up a new range of systems, capable of providing more intelligent functionality thank to information and communication (ICT). Networked embedded systems have revolutionized several application domains: energy networks, industrial automation and transport systems.

20th-century science and technology brought us effective methods and tools for designing both computational and physical systems. But the design of cyber-physical systems (CPS) is much more than the union of those two fields. Traditionally, information scientists only have a hazy notion of requirements imposed by the physical environment of computers. Similarly, mechanical, civil, and chemical engineers view computers strictly as devices executing algorithms. To the extent we have designed CPS, we have done so in an ad hoc, on-off manner that is not repeatable. A new science of CPS design will allow us to create new machines with complex dynamics and high reliability, to apply its principles to new industries and applications in a reliable and economically efficient way. Progress requires nothing less than the construction of a new science and technology foundation for CPS that is simultaneously physical and computational.

2.3. Motivations

Beyond the buzzword, a CPS is an ubiquitous object of our everyday life. CPSs have evolved from individual independent units (e.g an ABS brake) to more and more integrated networks of units, which may be aggregated into larger components or sub-systems. For example, a transportation monitoring network aggregates monitored stations and trains through a large scale distributed system with relatively high latency. Each individual train is being controlled by a train control network, each car in the train has its own real-time bus to control embedded devices. More and more, CPSs are mixing real-time low latency technology with higher latency distributed computing technology.

In the past 15 years, CPS development has moved towards Model Driven Engineering (MDE). With MDE methodology, first all requirements are gathered together with use cases, then a model of the system is built (sometimes several models) that satisfy the requirements. There are several modeling formalisms that have appeared in the past ten years with more or less success. The most successful are the *executable* models [0][0][0], i.e., models that can be simulated, exercised, tested and validated. This approach can be used for both software and hardware.

A common feature found in CPSs is the ever presence of concurrency and parallelism in models. Large systems are increasingly mixing both types of concurrency. They are structured hierarchically and comprise multiple synchronous devices connected by buses or networks that communicate asynchronously. This led to the advent of so-called GALS (Globally Asynchronous, Locally Synchronous) models, or PALS (Physically Asynchronous, Logically Synchronous) systems, where reactive synchronous objects are communicating asynchronously. Still, these infrastructures, together with their programming models, share some fundamental concerns: parallelism and concurrency synchronization, determinism and functional correctness, scheduling optimality and calculation time predictability.

Additionally, CPSs monitor and control real-world processes, the dynamics of which are usually governed by physical laws. These laws are expressed by physicists as mathematical equations and formulas. Discrete CPS models cannot ignore these dynamics, but whereas the equations express the continuous behavior usually using real numbers (irrational) variables, the models usually have to work with discrete time and approximate floating point variables.

[0] *Matlab/Simulink*, https://fr.mathworks.com/products/simulink.html
[0] *Ptolemy*, http://ptolemy.eecs.berkeley.edu
[0] *SysML*, http://www.uml-sysml.org

2.4. Challenges

A cyber-physical, or reactive, or embedded system is the integration of heterogeneous components originating from several design viewpoints: reactive software, some of which is embedded in hardware, interfaced with the physical environment through mechanical parts. Time takes different forms when observed from each of these viewpoints: it is discrete and event-based in software, discrete and time-triggered in hardware, continuous in mechanics or physics. Design of CPS often benefits from concepts of multiform and logical time(s) for their natural description. High-level formalisms used to model software, hardware and physics additionally alter this perception of time quite significantly.

In model-based system design, time is usually abstracted to serve the purpose of one of many design tasks: verification, simulation, profiling, performance analysis, scheduling analysis, parallelization, distribution, or virtual prototyping. For example in non-real-time commodity software, timing abstraction such as number of instructions and algorithmic complexity is sufficient: software will run the same on different machines, except slower or faster. Alternatively, in cyber-physical systems, multiple recurring instances of meaningful events may create as many dedicated logical clocks, on which to ground modeling and design practices.

Time abstraction increases efficiency in event-driven simulation or execution (i.e SystemC simulation models try to abstract time, from cycle-accurate to approximate-time, and to loosely-time), while attempting to retain functionality, but without any actual guarantee of valid accuracy (responsibility is left to the model designer). Functional determinism (a.k.a. conflict-freeness in Petri Nets, monotonicity in Kahn PNs, confluence in Milner's CCS, latency-insensitivity and elasticity in circuit design) allows for reducing to some amount the problem to that of many schedules of a single self-timed behavior, and time in many systems studies is partitioned into models of computation and communication (MoCCs). Multiple, multiform time(s) raises the question of combination, abstraction or refinement between distinct time bases. The question of combining continuous time with discrete logical time calls for proper discretization in simulation and implementation. While timed reasoning takes multiple forms, there is no unified foundation to reasoning about multi-form time in system design.

The objective of project-team TEA is henceforth to define formal models for timed quantitative reasoning, composition, and integration in embedded system design. Formal time models and calculi should allow us to revisit common domain problems in real-time system design, such as time predictability and determinism, memory resources predictability, real-time scheduling, mixed-criticality and power management; yet from the perspective gained from inter-domain timed and quantitative abstraction or refinement relations. A regained focus on fundamentals will allow to deliver better tooled methodologies for virtual prototyping and integration of embedded architectures.

3. Research Program

3.1. Previous Works

The challenges of team TEA support the claim that sound Cyber-Physical System design (including embedded, reactive, and concurrent systems altogether) should consider multi-form time models as a central aspect. In this aim, architectural specifications found in software engineering are a natural focal point to start from. Architecture descriptions organize a system model into manageable components, establish clear interfaces between them, collect domain-specific constraints and properties to help correct integration of components during system design. The definition of a formal design methodology to support heterogeneous or multi-form models of time in architecture descriptions demands the elaboration of sound mathematical foundations and the development of formal calculi and methods to instrument them. This constitutes the research program of team TEA.

System design based on the “synchronous paradigm” has focused the attention of many academic and industrial actors on abstracting non-functional implementation details from system design. This elegant design abstraction focuses on the logic of interaction in reactive programs rather than their timed behavior, allowing to secure functional correctness while remaining an intuitive programming model for embedded systems. Yet, it corresponds to embedded technologies of single cores and synchronous buses from the 90s, and may hardly cover the semantic diversity of distribution, parallelism, heterogeneity, of cyber-physical systems found in 21st century Internet-connected, true-time^{TM}-synchronized clouds, of tomorrow’s grids.

By contrast with a synchronous hypothesis yet from the same era, the polychronous MoCC implemented in the data-flow specification language Signal, available in the Eclipse project POP [0] and in the CCSL standard. [0], are inherently capable of describing multi-clock abstractions of GALS systems. The POP and TimeSquare projects provide tooled infrastructures to refine high-level specifications into real-time streaming applications or locally synchronous and globally asynchronous systems, through a series of model analysis, verification, and synthesis services. These tool-supported refinement and transformation techniques can assist the system engineer from the earliest design stages of requirement specification to the latest stages of synthesis, scheduling and deployment. These characteristics make polychrony much closer to the required semantic for compositional, refinement-based, architecture-driven, system design.

While polychrony was a step ahead of the traditional synchronous hypothesis, CCSL is a leap forward from synchrony and polychrony. The essence of CCSL is “multi-form time” toward addressing all of the domain-specific physical, electronic and logical aspects of cyber-physical system design.

3.2. Modeling Times

To make a sense and eventually formalize the semantics of time in system design, we should most certainly rely on algebraic representations of time found in previous works and introduce the paradigm of "time systems" (type systems to represent time) in a way reminiscent to CCSL. Just as a type system abstracts data carried along operations in a program, a time system abstracts the causal interaction of that program module or hardware element with its environment, its pre and post conditions, its assumptions and guarantees, either logical or numerical, discrete or continuous. Some fundamental concepts of the time systems we envision are present in the clock calculi found in data-flow synchronous languages like Signal or Lustre, yet bound to a particular model of concurrency, hence time.

In particular, the principle of refinement type systems [0], is to associate information (data-types) inferred from programs and models with properties pertaining, for instance, to the algebraic domain on their value, or any algebraic property related to its computation: effect, memory usage, pre-post condition, value-range, cost, speed, time, temporal logic [0]. Being grounded on type and domain theories, a time system should naturally be equipped with program analysis techniques based on type inference (for data-type inference) or abstract interpretation (for program properties inference) to help establish formal relations between heterogeneous component “types”. Just as a time calculus may formally abstract timed concurrent behaviors of system components, timed relations (abstraction and refinement) represent interaction among components.

Scalability and compositionality requires the use of assume-guarantee reasoning to represent them, and to facilitate composition by behavioral sub-typing, in the spirit of the (static) contract-based formalism proposed by Passerone et al. [0]. Verification problems encompassing heterogeneously timed specifications are common and of great variety: checking correctness between abstract and concrete time models relates to desynchronisation (from synchrony to asynchrony) and scheduling analysis (from synchrony to hardware). More generally, they can be perceived from heterogeneous timing viewpoints (e.g. mapping a synchronous-time software on a real-time middle-ware or hardware).

[0] *Polychrony on Polarsys*, https://www.polarsys.org/projects/polarsys.pop
[0] *Clock Constraints in UML/MARTE CCSL*. C. André, F. Mallet. RR-6540. Inria, 2008. http://hal.inria.fr/inria-00280941
[0] *Abstract Refinement Types*. N. Vazou, P. Rondon, and R. Jhala. European Symposium on Programming. Springer, 2013.
[0] *LTL types FRP*. A. Jeffrey. Programming Languages meets Program Verification.
[0] *A contract-based formalism for the specification of heterogeneous systems*. L. Benvenistu, et al. FDL, 2008

This perspective demands capabilities not only to inject time models one into the other (by abstract interpretation, using refinement calculi), to compare time abstractions one another (using simulation, refinement, bi-simulation, equivalence relations) but also to prove more specific properties (synchronization, determinism, endochrony). All this formalization effort will allow to effectively perform the tooled validation of common cross-domain properties (e.g. cost v.s. power v.s. performance v.s. software mapping) and tackle equally common yet though case studies such as these linking battery capacity, to on-board CPU performance, to static software schedulability, to logical software correctness and plant controllability: the choice of the right sampling period across the system components.

3.3. Modeling Architectures

To address the formalization of such cross-domain case studies, modeling the architecture formally plays an essential role. An architectural model represents components in a distributed system as boxes with well-defined interfaces, connections between ports on component interfaces, and specifies component properties that can be used in analytical reasoning about the model. Several architectural modeling languages for embedded systems have emerged in recent years, including the SAE AADL [0], SysML [0], UML MARTE [0].

In system design, an architectural specification serves several important purposes. First, it breaks down a system model into manageable components to establish clear interfaces between components. In this way, complexity becomes manageable by hiding details that are not relevant at a given level of abstraction. Clear, formally defined, component interfaces allow us to avoid integration problems at the implementation phase. Connections between components, which specify how components affect each other, help propagate the effects of a change in one component to the linked components.

Most importantly, an architectural model is a repository to share knowledge about the system being designed. This knowledge can be represented as requirements, design artifacts, component implementations, held together by a structural backbone. Such a repository enables automatic generation of analytical models for different aspects of the system, such as timing, reliability, security, performance, energy, etc. Since all the models are generated from the same source, the consistency of assumptions w.r.t. guarantees, of abstractions w.r.t. refinements, used for different analyses becomes easier, and can be properly ensured in a design methodology based on formal verification and synthesis methods.

Related works in this aim, and closer in spirit to our approach (to focus on modeling time) are domain-specific languages such as Prelude [0] to model the real-time characteristics of embedded software architectures. Conversely, standard architecture description languages could be based on algebraic modeling tools, such as interface theories with the ECDAR tool [0].

In project TEA, it takes form by the normalization of the AADL standard's formal semantics and the proposal of a time specification annex in the form of related standards, such as CCSL, to model concurrency time and physical properties, and PSL, to model timed traces.

3.4. Scheduling Theory

Based on sound formalization of time and CPS architectures, real-time scheduling theory provides tools for predicting the timing behavior of a CPS which consists of many interacting software and hardware components. Expressing parallelism among software components is a crucial aspect of the design process of a CPS. It allows for efficient partition and exploitation of available resources.

The literature about real-time scheduling [0] provides very mature schedulability tests regarding many scheduling strategies, preemptive or non-preemptive scheduling, uniprocessor or multiprocessor scheduling, etc. Scheduling of data-flow graphs has also been extensively studied in the past decades.

[0] *Architecture Analysis and Design Language*, AS-5506. SAE, 2004. http://standards.sae.org/as5506b

[0] *System modeling Language*. OMG, 2007. http://www.omg.org/spec/SysML

[0] *UML Profile for MARTE*. OMG, 2009. http://www.omg.org/spec/MARTE

[0] *The Prelude language*. LIFL and ONERA, 2012. http://www.lifl.fr/~forget/prelude.html

[0] *PyECDAR, timed games for timed specifications*. Inria, 2013. https://project.inria.fr/pyecdar

[0] *A survey of hard real-time scheduling for multiprocessor systems*. R. I. Davis and A. Burns. *ACM Computing Survey* 43(4), 2011.

A milestone in this prospect is the development of abstract affine scheduling techniques [0]. It consists, first, of approximating task communication patterns (here Safety-Critical Java threads) using cyclo-static data-flow graphs and affine functions. Then, it uses state of the art ILP techniques to find optimal schedules and concretize them as real-time schedules for Safety Critical Java programs [00].

Abstract scheduling, or the use of abstraction and refinement techniques in scheduling borrowed to the theory of abstract interpretation [0] is a promising development toward tooled methodologies to orchestrate thousands of heterogeneous hardware/software blocks on modern CPS architectures (just consider modern cars or aircrafts). It is an issue that simply defies the state of the art and known bounds of complexity theory in the field, and consequently requires a particular address.

To develop the underlying theory of this promising research topic, we first need to deepen the theoretical foundation to establish links between scheduling analysis and abstract interpretation. A theory of time systems would offer the ideal framework to pursue this development. It amounts to representing scheduling constraints, inferred from programs, as types or contract properties. It allows to formalize the target time model of the scheduler (the architecture, its middle-ware, its real-time system) and defines the basic concepts to verify assumptions made in one with promises offered by the other: contract verification or, in this case, synthesis.

3.5. Virtual Prototyping

Virtual Prototyping is the technology of developing realistic simulators from models of a system under design; that is, an emulated device that captures most, if not all, of the required properties of the real system, based on its specifications. A virtual prototype should be run and tested like the real device. Ideally, the real application software would be run on the virtual prototyping platform and produce the same results as the real device with the same sequence of outputs and reported performance measurements. This may be true to some extent only. Some trade-offs have often to be made between the accuracy of the virtual prototype, and time to develop accurate models.

In order to speed-up simulation time, the virtual prototype must trade-off with something. Depending upon the application designer's goals, one may be interested in trading some loss of accuracy in exchange for simulation speed, which leads to constructing simulation models that focus on some design aspects and provide abstraction of others. A simulation model can provide an abstraction of the simulated hardware in three directions:

- *Computation abstraction*. A hardware component computes a high level function by carrying out a series of small steps executed by composing logical gates. In a virtual prototyping environment, it is often possible to compute the high level function directly by using the available computing resources on the simulation host machine, thus abstracting the hardware function.
- *Communication abstraction*. Hardware components communicate together using some wiring, and some protocol to transmit the data. Simulation of the communication and the particular protocol may be irrelevant for the purpose of virtual prototyping: communication can be abstracted into higher level data transmission functions.
- *Timing Abstraction*. In a cycle accurate simulator, there are multiple simulation tasks, and each task makes some progress on each clock cycle, but this slows down the simulation. In a virtual prototyping experiment, one may not need such precise timing information: coarser time abstractions can be defined allowing for faster simulation.

The cornerstone of a virtual prototyping platform is the component that simulates the processor(s) of the platform, and its associated peripherals. Such simulation can be *static* or *dynamic*.

[0] *Buffer minimization in EDF scheduling of data-flow graphs*. A. Bouakaz and J.-P. Talpin. LCTES, ACM, 2013.
[0] *ADFG for the synthesis of hard real-time applications*. A. Bouakaz, J.-P. Talpin, J. Vitek. ACSD, IEEE, June 2012.
[0] *Design of SCJ Level 1 Applications Using Affine Abstract Clocks*. A. Bouakaz and J.-P. Talpin. SCOPES, ACM, 2013.
[0] *La vérification de programmes par interprétation abstraite*. P. Cousot. Séminaire au Collège de France, 2008.

A solution usually adopted to handle time in virtual prototyping is to manage hierarchical time scales, use component abstractions where possible to gain performance, use refinement to gain accuracy where needed. Localized time abstraction may not only yield faster simulation, but facilitate also verification and synthesis (e.g. synchronous abstractions of physically distributed systems). Such an approach requires computations and communications to be harmoniously discretized and abstracted from originally heterogeneous viewpoints onto a structuring, articulating, pivot model, for concerted reasoning about time and scheduling of events in a way that ensures global system specification correctness.

In the short term these component models could be based on libraries of predefined models of different levels of abstractions. Such abstractions are common in large programming workbench for hardware modeling, such as SystemC, but less so, because of the engineering required, for virtual prototyping platforms.

The approach of team TEA provides an additional ingredient in the form of rich component interfaces. It therefore dictates to further investigate the combined use of conventional virtual prototyping libraries, defined as executable abstractions of real hardware, with executable component simulators synthesised from rich interface specifications (using, e.g., conventional compiling techniques used for synchronous programs).

4. Application Domains

4.1. Automotive and Avionics

From our continuous collaboration with major academic and industrial partners through projects TOPCASED, OPENEMBEDD, SPACIFY, CESAR, OPEES, P and CORAIL, our experience has primarily focused on the aerospace domain. The topics of time and architecture of team TEA extend to both avionics and automotive. Yet, the research focus on time in team TEA is central in any aspect of, cyber-physical, embedded system design in factory automation, automotive, music synthesis, signal processing, software radio, circuit and system on a chip design; many application domains which, should more collaborators join the team, would definitely be worth investigating.

Multi-scale, multi-aspect time modeling, analysis and software synthesis will greatly contribute to architecture modeling in these domains, with applications to optimized (distributed, parallel, multi-core) code generation for avionics (project Corail with Thales avionics, section 8) as well as modeling standards, real-time simulation and virtual integration in automotive (project with Toyota ITC, section 8).

Together with the importance of open-source software, one of these projects, the FUI Project P (section 8), demonstrated that a centralized model for system design could not just be a domain-specific programming language, such as discrete Simulink data-flows or a synchronous language. Synchronous languages implement a fixed model of time using logical clocks that are abstraction of time as sensed by software. They correspond to a fixed viewpoint in system design, and in a fixed hardware location in the system, which is not adequate to our purpose and must be extended.

In project P, we first tried to define a centralized model for importing discrete-continuous models onto a simplified implementation of SIMULINK: P models. Certified code generators would then be developed from that format. Because this does not encompass all aspects being translated to P, the P meta-model is now being extended to architecture description concepts (of the AADL) in order to become better suited for the purpose of system design. Another example is the development of System modeler on top of SCADE, which uses the more model-engineering flavored formalism SysML to try to unambiguously represent architectures around SCADE modules.

An abstract specification formalism, capable of representing time, timing relations, with which heterogeneous models can be abstracted, from which programs can be synthesized, naturally appears better suited for the purpose of virtual prototyping. RT-Builder, based on Signal like Polychrony and developed by TNI, was industrially proven and deployed for that purpose at Peugeot. It served to develop the virtual platform simulating all on-board electronics of PSA cars. This ‘hardware in the loop” simulator was used to test equipments supplied by other manufacturers with respect to virtual cars. In the advent of the related automotive standard, RT-Builder then became AUTOSAR-Builder.

4.2. Factory Automation

In collaboration with Mitsubishi R&D, we explore another application domain where time and domain heterogeneity are prime concerns: factory automation. In factory automation alone, a system is conventionally built from generic computing modules: PLCs (Programmable Logic Controllers), connected to the environment with actuators and detectors, and linked to a distributed network. Each individual, physically distributed, PLC module must be timely programmed to perform individually coherent actions and fulfill the global physical, chemical, safety, power efficiency, performance and latency requirements of the whole production chain. Factory chains are subject to global and heterogeneous (physical, electronic, functional) requirements whose enforcement must be orchestrated for all individual components.

Model-based analysis in factory automation emerges from different scientific domains and focus on different CPS abstractions that interact in subtle ways: logic of PLC programs, real-time electromechanical processing, physical and chemical environments. This yields domain communication problems that render individual domain analysis useless. For instance, if one domain analysis (e.g. software) modifies a system model in a way that violates assumptions made by another domain (e.g. chemistry) then the detection of its violation may well be impossible to explain to either of the software and chemistry experts. As a consequence, cross-domain analysis issues are discovered very late during system integration and lead to costly fixes. This is particularly prevalent in multi-tier industries, such as avionic, automotive, factories, where systems are prominently integrated from independently-developed parts.

5. Highlights of the Year

5.1. Highlights of the Year

In 2016, TEA was successfully evaluated, one year after its creation. The team started fruitful collaborations with UC San Diego, with Mitsubishi R&D, with ASTRI, to elaborate our research program on system composition, verification, and simulation toward novel applications perspectives in codesign, operating system design, factory automation, robotics.

6. New Software and Platforms

6.1. ADFG: Affine data-flow graphs scheduler synthesis

Participants: Alexandre Honorat, Jean-Pierre Talpin, Thierry Gautier, Loïc Besnard.

We proposed [2], and implemented [0], a new data-flow design model: ADFG, initially to synthesize schedulers for SCJ/L1 applications. The principle of ADFG is to perform a linear abstraction of complex cyclo-static scheduling problems followed by the exploration of a concrete solution extracted from the abstract solution space, hence the name: abstract affine data-flow scheduling. ADFG guarantees schedules that ordinary (e.g. RTJ, SCJ) task-sets do not cause overflows or underflows. ADFG objectives are to maximize the throughput (the processors utilization) while minimizing buffering storage space needed between actors. ADFG supports EDF and fixed-priority scheduling policies for uni-, multi-processors and distributed systems.

The data-flow design model of ADFG comes with a development tool integrated in the Eclipse IDE for easing the development of SCJ/L1 applications and enforcing the restrictions imposed by the design model. It consists of a GMF editor where applications are designed graphically and timing and buffering parameters can be synthesized. Abstract affine scheduling is first applied on the data-flow subgraph, that consists only of periodic actors, to compute timeless scheduling constraints (e.g. relation between the speeds of two actors) and buffering parameters. Then, symbolic fixed-priority schedulability analysis (i.e., synthesis of timing and scheduling parameters of actors) considers both periodic and aperiodic actors.

[0] *The ADFG tool*, Adnan Bouakaz, http://people.irisa.fr/Adnan.Bouakaz/software.htm

In the case of safety-critical Java, and through a model-to-text transformations using Acceleo, SCJ code for missions, interfaces of handlers, and the mission sequencer is automatically generated in addition to the annotations needed by the memory checker. Channels are implemented as cyclic arrays or cyclical asynchronous buffers; and a fixed amount of memory is hence reused to store the infinite streams of tokens.

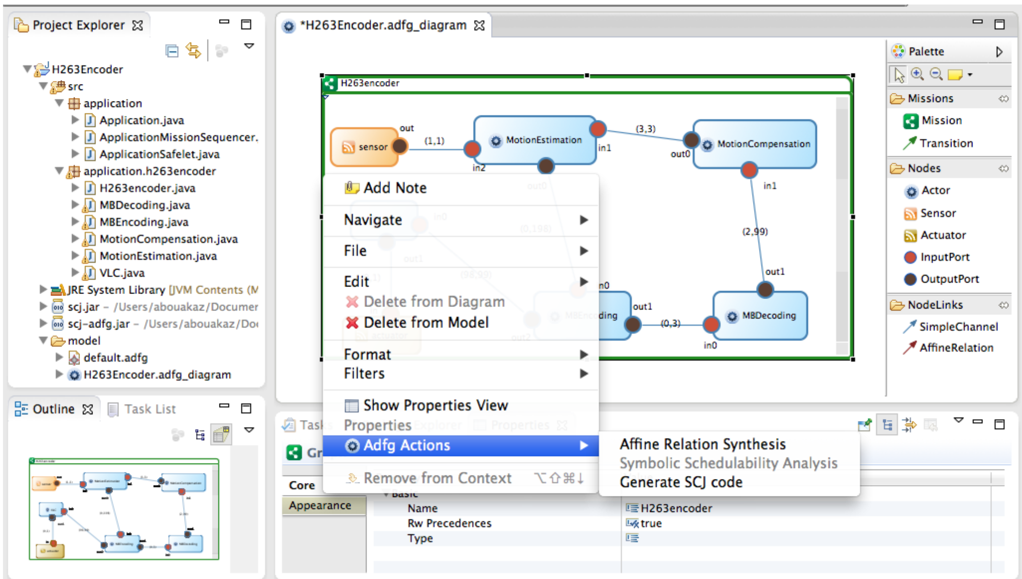


Figure 1. The ADFG tool

6.2. The Eclipse project POP

Participants: Loïc Besnard, Thierry Gautier, Jean-Pierre Talpin.

The distribution of project POP is a major achievement of the ESPRESSO (and now TEA) project-team. The Eclipse project POP is a model-driven engineering front-end to our open-source toolset Polychrony. It was finalized in the frame of project OPEES, as a case study: by passing the POLARSYS qualification kit as a computer aided simulation and verification tool. This qualification was implemented by CS Toulouse in conformance with relevant generic (platform independent) qualification documents. Polychrony is now distributed by the Eclipse project POP on the platform of the POLARSYS industrial working group. Project-team TEA aims at continuing its dissemination to academic partners, as to its principles and features, and industrial partners, as to the services it can offer.

Project POP is composed of the Polychrony tool set, under GPL license, and its Eclipse framework, under EPL license. SSME (Syntactic Signal-Meta under Eclipse), is the meta-model of the Signal language implemented with Eclipse/Ecore. It describes all syntactic elements specified in Signal Reference Manual [0]: all Signal operators (e.g. arithmetic, clock synchronization), model (e.g. process frame, module), and construction (e.g. iteration, type declaration). The meta-model primarily aims at making the language and services of the Polychrony environment available to inter-operation and composition with other components (e.g. AADL, Simulink, GeneAuto, P) within an Eclipse-based development tool-chain. Polychrony now comprises the

0

SIGNAL V4-Inria version: Reference Manual. Besnard, L., Gautier, T. and Le Guernic, P. http://www.irisa.fr/espresso/Polychrony, 2010

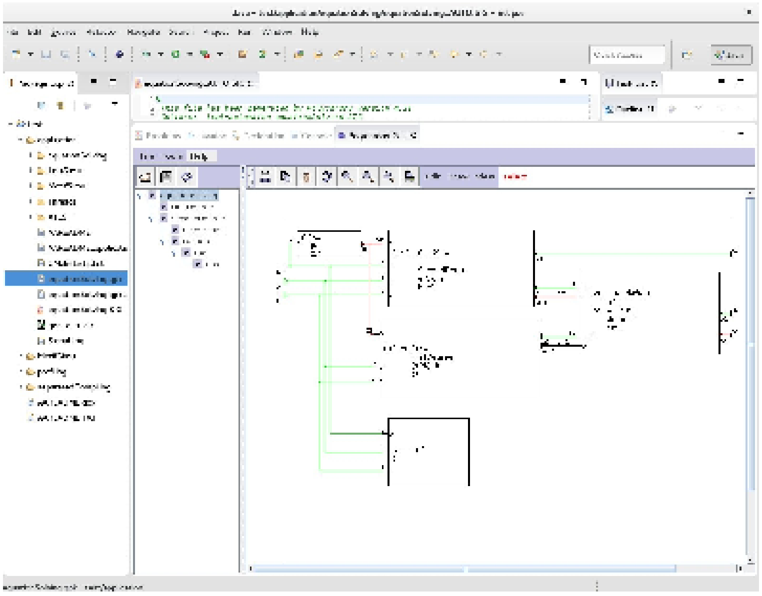


Figure 2. The Eclipse POP Environment

capability to directly import and export Ecore models instead of textual Signal programs, in order to facilitate interaction between components within such a tool-chain. The download site for project POP has opened in 2015 at https://www.polarsys.org/projects/polarsys.pop. It should be noted that the Eclipse Foundation does not host code under GPL license. So, the Signal toolbox useful to compile Signal code from Eclipse is hosted on our web server.

6.3. The Polychrony toolset

Participants: Loïc Besnard, Thierry Gautier, Jean-Pierre Talpin.

The Polychrony toolset is an Open Source development environment for critical/embedded systems. It is based on Signal, a real-time polychronous data-flow language. It provides a unified model-driven environment to perform design exploration by using top-down and bottom-up design methodologies formally supported by design model transformations from specification to implementation and from synchrony to asynchrony. It can be included in heterogeneous design systems with various input formalisms and output languages. The Polychrony tool-set provides a formal framework to: validate a design at different levels, by the way of formal verification and/or simulation; refine descriptions in a top-down approach; abstract properties needed for black-box composition; compose heterogeneous components (bottom-up with COTS); generate executable code for various architectures. The Polychrony tool-set contains three main components and an experimental interface to GNU Compiler Collection (GCC):

- The Signal toolbox, a batch compiler for the Signal language, and a structured API that provides a set of program transformations. Itcan be installed without other components and is distributed under GPL V2 license.
- The Signal GUI, a Graphical User Interface to the Signal toolbox (editor + interactive access to compiling functionalities). It can be used either as a specific tool or as a graphical view under Eclipse. In 2015, it has been transformed and restructured, in order to get a more up-to-date interface allowing

multi-window manipulation of programs. It is distributed under GPL V2 license.

- The SSME platform, a front-end to the Signal toolbox in the Eclipse environment. It is distributed under EPL license.

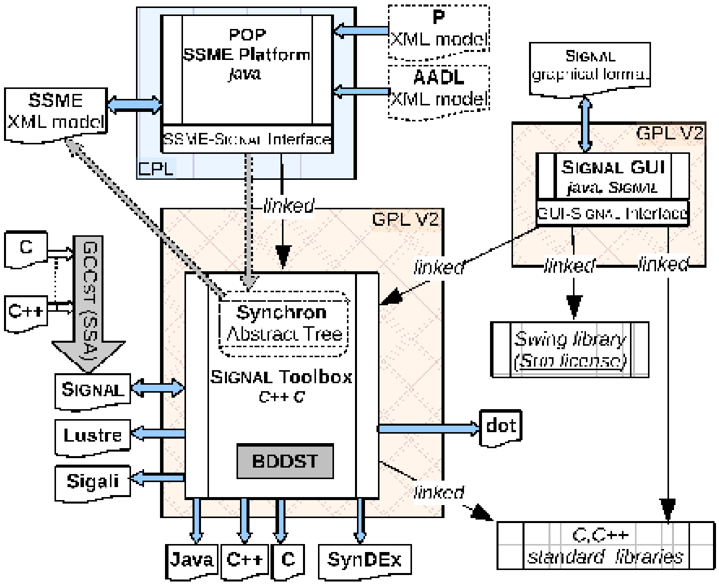


Figure 3. The Polychrony toolset high-level architecture

As part of its open-source release, the Polychrony tool-set not only comprises source code libraries but also an important corpus of structured documentation, whose aim is not only to document each functionality and service, but also to help a potential developer to package a subset of these functionalities and services, and adapt them to developing a new application-specific tool: a new language front-end, a new back-end compiler. This multi-scale, multi-purpose documentation aims to provide different views of the software, from a high-level structural view to low-level descriptions of basic modules. It supports a distribution of the software "by apartment" (a functionality or a set of functionalities) intended for developers who would only be interested by part of the services of the tool-set. The Polychrony tool-set also provides a large library of Signal programs and examples, user documentations and developer-oriented implementation documents, and facilities to generate new versions. The Polychrony tool-set can be freely downloaded from http://polychrony.inria.fr/. This site, intended for users and for developers, contains executable and source versions of the software for different platforms, user documentation, examples, libraries, scientific publications and implementation documentation. In particular, this is the site for the open-source distribution of Polychrony. The Inria GForge https://gforge.inria.fr contains the whole source of the environment and its documentation. It is intended for developers.

7. New Results

7.1. Toward a distribution of ADFG

Participants: Alexandre Honorat, Jean-Pierre Talpin, Thierry Gautier, Loïc Besnard.

The ADFG tool is being developed in the context of the ADT "Opama" in order to serve both as scheduler synthesis tool from AADL specifications and ordinary tasksets. ADFG has been partly rewritten in order to target more users : it is now freely available online and comes with a complete documentation. These improvements imply that ADFG does not anymore provide Safety Critical Java application generation; its main purpose of scheduler synthesis is now available from an Eclipse plugin, as a command-line interface, and also in Polychrony (as part of the AADL to Signal translation process). Moreover ADFG accepts and exports several file formats with related scheduling tools: SDF3, Yartiss and soon Cheddar.

The Eclipse interface has changed significantly with a dialog window and a console to present the results (as shown in the figure 4). Also the graphical data-flow graph editor is still present but has been simplified. An other big change (not seen by the end-user) is the internal use of the free LpSolve linear programming software instead of CPLEX. Finally, it will soon be possible to use this software not only as a scheduling synthesizer but also as a scheduling checker (with timing properties given by the user).

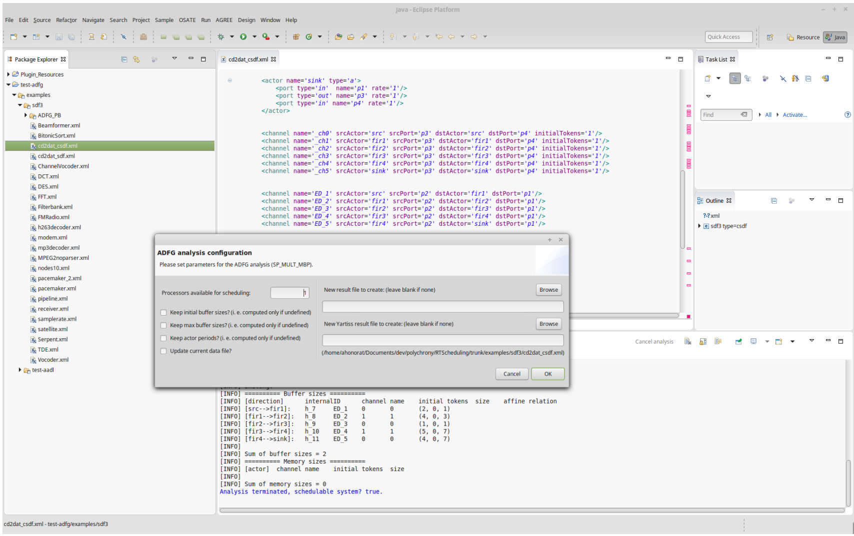


Figure 4. ADFG under Eclipse

7.2. Modular verification of cyber-physical systems using contract theory

Participants: Jean-Pierre Talpin, Benoit Boyer, David Mentre, Simon Lunel.

The primary goal of our project, in collaboration with Mitsubishi Electronics Research Centre Europe (MERCE), is to ensure correctness-by-design in realistic cyber-physical systems, i.e., systems that mix software and hardware in a physical environment, e.g., Mitsubishi factory automation lines or water-plant factory. To achieve that, we develop a verification methodology based on contract reasoning.

We have first performed a state of the art of the research and the work of A. Platzer with the Differential Dynamic Logic ($d\mathcal{L}$) retained our attention [0]. This a formalism built on the Dynamic Logic of V. Pratt augmented with the possibility of expressing Ordinary Differential Equations (ODEs). ODEs are the usual way to model physical behaviors in physics and $d\mathcal{L}$ permits to accurately model cyber-physical systems. But this logic can also express properties on real arithmetic and there is proof system associated, under the form of a sequent calculus, which let us a mean to prove specifications. To finish, it is very natural to use contract to specify systems since it was the primary goal of the work of V. Pratt. To conclude, $d\mathcal{L}$ is particularly fit to our purpose.

[0] *Differential Dynamic Logic for Hybrid Systems*, André Platzer, http://symbolaris.com/logic/dL.html

We have some preliminary results about a design-by-composition methodology: we have defined a syntactic composition operator in $d\mathcal{L}$, which enjoys associativity and commutativity. We have then characterized the conditions under which we can derive automatically a proof of the contract of our composition. To exemplified our ideas, we are currently studying a simplified water-tank system, which will serve as a basis for more realistic case studies. We plan to provide refinement and abstraction mechanisms to ultimately allow a mix between vertical and horizontal design.

7.3. Runtime verification and trace analysis

Participants: Vania Joloboff, Daian Yue, Frédéric Mallet.

When engineers design a new cyber physical system, there are well known requirements that can be translated as system properties that must be verified. These properties can be expressed in some formalism and when the model has been designed, the properties can be checked at the model level, using model checking techniques or other model verification techniques.

This requires that the properties are well specified at the time the virtual prototype is assembled. However it is also the case that many intrinsic properties are actually unforeseen when the virtual prototype is assembled, for example that some hardware buffer overflow should not remain unnoticed by the software. In most cases, during system design the simulation fails: the engineers then must investigate the cause of the failure.

A widely used technique for that consists in storing all of the trace data of simulation sessions into trace files, which are analyzed later with specialized trace analyzer tools. Such trace files have become huge, possibly hundred of Gigabytes as all data are stored into the trace files, and have become intractable by human manual handling.

In order to better identify the reason for such failures and capture the missing properties that the system should verify we have started to work on a new run time verification approach based on trace analysis. Approaches like PSL requires that the properties to verify are known before hand. Our approach is attempting for the engineers to experiment various property verification of failing simulations without re-building the virtual prototype. We have established a technique that makes it possible to investigate properties either statically working from a trace file or dynamically by introducing a dynamic verification component into the virtual prototype, or actually the real system.

The Trace Runtime Analysis Platform (TRAP) provides a model-based framework and implements the corresponding tool chain to support runtime analysis and verification of traces generated by virtual prototypes or cyber-physical systems. The main goal is to make it easy for engineers to define system properties that should be satisfied and verify them at system runtime (or from a recorded session). The property verification tools proposed do not require a detailed knowledge of the system implementation, do not require any modification or recompilation of the system to investigate different properties, and do not require the engineers to be familiar with temporal logic. TRAP proposes Domain Specific Languages (DSL's) integrated within the Eclipse Modeling Framework to express the properties. The DSL tool-chain uses the concept of Logical Clock defined by CCSL and takes advantage of CCSL clock algebra as the underlying formal support. The DSL's compilers eventually generate C++ code to verify the properties at run time, making usage of dynamically loaded code.

This year we have investigated and implemented this approach, using Eclipse EMF. The STML and TPSL compilers are implemented in Java and generate C++ code. Results of this work have been published at the FDL'16 conference referenced on IEEE Explore. [17]

7.4. Polychronous automata and formal validation of AADL models

Participants: Loïc Besnard, Thierry Gautier, Alexandre Honorat, Clément Guy, Jean-Pierre Talpin.

We have defined a model of *polychronous automata* based on clock relations [7]. A specificity of this model is that an automaton is submitted to clock constraints: these finite-state automata define transition systems to express explicit reactions together with properties, in the form of Boolean formulas over logical time, to constrain their behavior. This allows one to specify a wide range of control-related configurations, either reactive, or restrictive with respect to their control environment. A semantic model is defined for these polychronous automata, that relies on a Boolean algebra of clocks. Polychronous automata integrate smoothly with data-flow equations in the polychronous model of computation.

This polychronous MoC has been used previously as semantic model for systems described in the core AADL standard. The core AADL is extended with annexes, such as the Behavior Annex, which allows to specify more precisely architectural behaviors. The translation from AADL specifications into the polychronous model should take into account these behavior specifications, which are based on description of automata.

For that purpose, the AADL state transition systems are translated as Signal automata (a slight extension of the Signal language has been defined to support the model of polychronous automata). States are declared as Signal labels. Transitions are expressed using a call to a specific Signal process `Automaton\_Transition` which takes as parameters the labels of the source and destination states, and the condition expression corresponding to the AADL guard of the transition. The transition processes implicitly declare the equations that are required to compute the firing instants of the transitions. These processes, viewed as macros, are replaced during Signal compilation with a set of Signal equations handling current state and transition firing.

Once the AADL model of a system transformed into a Signal program, one can analyze the program using the Polychrony framework in order to check if timing, scheduling and logical requirements over the whole system are met.

We have implemented the translation and experimented it using a concrete case study, which is the AADL modeling of an Adaptive Cruise Control (ACC) system, a highly safety-critical system embedded in recent cars.

7.5. Formal Semantics of Behavior Specifications in the Architecture Analysis and Design Language Standard

Participants: Loïc Besnard, Thierry Gautier, Clément Guy, Jean-Pierre Talpin.

In system design, an architecture specification or model serves, among other purposes, as a repository to share knowledge about the system being designed. Such a repository enables automatic generation of analytical models for different aspects relevant to system design (timing, reliability, security, etc.). The Architecture Analysis and Design Language (AADL) is a standard proposed by SAE to express architecture specifications and share knowledge between the different stakeholders about the system being designed. To support unambiguous reasoning, formal verification, high-fidelity simulation of architecture specifications in a model-based AADL design work-flow, we have defined a formal semantics for the behavior specification of the AADL. Since it began being discussed in the AADL standard committee, our formal semantics evolved from a synchronous model of computation and communication to a semantic framework for time and concurrency in the standard: asynchronous, synchronous or timed, to serve as a reference for model checking, code generation or simulation tools uses with the standard [14]. These semantics are simple, relying on the structure of automata present in the standard already, yet provide tagged, trace semantics framework to establish formal relations between (synchronous, asynchronous, timed) usages or interpretations of behavior.

We define the model of computation and communication of a behavior specification by the synchronous, timed or asynchronous traces of automata with variables. These constrained automata are derived from *polychronous automata* defined within the polychronous model of computation and communication [7].

States of a behavior annex transition system can be either observable from the outside (`initial`, `final` or `complete` states), that is states in which the execution of the component is paused or stopped and its outputs are available; or non observable execution states, that is internal states. We thus define two kinds of steps in the transition system: *small steps*, that is non-observable steps from or to an internal state; and *big steps*, that is observable steps from a *complete* state to another, through a number of small steps). The semantics of the AADL considers the observable states of the automaton. The set of states S_A of automaton A (used to interpret the behavior annex) thus only contains states corresponding to these observable states and the set of transitions T_A big-step transitions from an observable state to another (by opposition with small-step transitions from or to an execution state). The action language of the behavior annex defines actions performed during transitions. Actions associated with transitions are action blocks that are built from basic actions and a minimal set of control structures (sequences, sets, conditionals and loops). Typically, a behavior action sequence is represented by concatenating the transition systems of its elements; a behavior action set is represented by composing the transition systems of its elements.

For our semantics, we considered a significant subset of the behavioral specification annex of the AADL. This annex allows one to attach a behavior specification to any components of a system modeled using the AADL, and can be then analyzed for different purposes which could be, for example, the verification of logical, timing or scheduling requirements.

7.6. Integration of Polychrony with QGen

Participants: Loïc Besnard, Thierry Gautier, Christophe Junke, Jean-Pierre Talpin.

The FUI project P gave birth to the GGen qualifiable model compiler, developed by Adacore. The tool accepts a discrete subset of Simulink expressed in a language called P and produces C or Ada code.

Our contribution was about providing a semantic bridge between Polychrony and QGen [15]. Our objective was to use Polychrony to compute fined-grained static scheduling of computations and communications for P models based on architectural properties. This work was twofold. First, we defined an alternative unambiguous static block scheduler for QGen, which can compute both partial and total orders based on user preferences. The purpose of this sequencer is to allow QGen to inter-operate with external sequencing tools while providing guarantees about the compatibility of external block execution orders with respect to both QGen's compilation scheme and user expectations. On the other hand, we developed a transformation function from the P language, more precisely, from the System Model subset of P, to the Signal meta-model, SSME. This work is based on a high-level API designed on top of SSME and can be used to transform a subset of Simulink to Signal. We validated our approach with the test suite used by QGen which is composed of over two-hundred small-sized Simulink models. We tested both block sequencing and model transformations. We ran the conversion tool and the set of models used by QGen for its regression tests and successfully converted medium to large models. The P language is capable of representing a useful subset of Simulink. That is why it is an interesting tool to help interpreting Simulink models and possibly architectural properties as executable Signal programs. The programs currently produced with our transformation tool can be compiled by Polychrony and reorganized as clusters of smaller processes.

7.7. Code generation for poly-endochronous processes

Participants: Loïc Besnard, Thierry Gautier, Jean-Pierre Talpin.

The synchronous modeling paradigm provides strong correctness guarantees for embedded system design while requiring minimal environmental assumptions. In most related frameworks, global execution correctness is achieved by ensuring the insensitivity of (logical) time in the program from (real) time in the environment. Tis property, called endochrony, can be statically checked, making it fast to ensure design correctness. Unfortunately, it is not preserved by composition, which makes it difficult to exploit with component-based design concepts in mind. It has been shown that compositionality can be achieved by weakening the objective of endochrony: a weakly endochronous system is a deterministic system that can perform independent computations and communications in any order as long as this does not alter its global state. Moreover, the non-blocking composition of weakly endochronous processes is isochronous, which means that the synchronous and asynchronous compositions of weakly endochronous processes accept the same behaviors. Unfortunately, testing weak endochrony needs state-space exploration, which is very costly in the general case. Then, a particular case of weak endochrony, called polyendochrony, was defined, which allows static checking thanks to the existing clock calculus. The clock hierarchy of a polyendochronous system may have several trees, with synchronization relations between clocks placed in different trees, but the clock expressions of the clock system must be such that there is no clock expression (especially, no root clock expression) defined by symmetric difference: root clocks cannot refer to absence. In other words, the clock system must be in disjunctive form [10].

We have now implemented code generation for poly-endochronous systems in Polychrony. This generation reuses techniques of distributed code generation, with rendez-vous management for synchronization constraints on clocks which are not placed in the same tree of clocks. The approach has been validated on several use cases running in parallel with time to time synchronization.

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Contracts with Industry

8.1.1. Toyota Info-Technology Centre (2014-2016)

Title: Co-Modeling of Safety-Critical Multi-threaded Embedded Software for Multi-Core Embedded Platforms

Inria principal investigator: Jean-Pierre Talpin

International Partner (Institution - Laboratory - Researcher):

Toyota Info-Technology Centre, Mountain View, California

Virginia Tech Research Laboratories, Arlington

Duration: renewed yearly since 2014

Abstract: We started a new project in April 2014 funded by Toyota ITC, California, to work with Huafeng Yu (a former post-doctorate of team ESPRESSO) and with VTRL as US partner. The main topic of our project is the semantic-based model integration of automotive architectures, virtual integration, toward formal verification and automated code synthesis. This year, Toyota ITC is sponsoring our submission for the standardization of a time annex in the SAE standard AADL.

In a second work-package, we aim at elaborating a standardized solution to virtually integrate and simulate a car based on heterogeneous models of its components. This year, it will be exemplified by the elaboration of a case study in collaboration with Virginia Tech. The second phase of the project will consist of delivering an open-source, reference implementation, of the proposed AADL standard and validate it with a real-scale model of the initial case-study.

8.2. Bilateral Grants with Industry

8.2.1. Mitsubishi Electric R&D Europe (2015-2018)

Title: Analysis and verification for correct by construction orchestration in automated factories

Inria principal investigator: Jean-Pierre Talpin, Simon Lunel

International Partner: Mitsubishi Electric R&D Europe

Duration: 2015 - 2018

Abstract: The primary goal of our project is to ensure correctness-by-design in cyber-physical systems, i.e., systems that mix software and hardware in a physical environment, e.g., Mitsubishi factory automation lines. We plan to explore a multi-sorted algebraic framework for static analysis and formal verification starting from a simple use case extracted from Mitsubishi factory automation documentations. This will serve as a basis to more ambitious research where we intend to leverage recent advance in type theory, SMT solvers for nonlinear real arithmetic (dReal and δ-decidability) and contracts theory (meta-theory of Benveniste et al., Ruchkin's contracts) to provide a general framework of reasoning about heterogeneous factory components.

9. Partnerships and Cooperations

9.1. National Initiatives

9.1.1. ANR

Program: ANR

Project acronym: **Feever**

Project title: Faust Environment Everyware

Duration: 2014-2016

Coordinator: Pierre Jouvelot, Mines ParisTech

Other partners: Grame, Inria Rennes, CIEREC

URL: http://www.feever.fr

Abstract:

The aim of project FEEVER is to ready the Faust music synthesis language for the Web. In this context, we collaborate with Mines ParisTech to define a type system suitable to model music signals timed at multiple rates and to formally support playing music synthesized from different physical locations.

9.1.2. PAI CORAC

Program: CORAC

Project acronym: CORAIL

Project title: Composants pour l'Avionique Modulaire Étendue

Duration: July 2013 - May 2017

Coordinator: Thales Avionics

Other partners: Airbus, Dassault Aviation, Eurocopter, Sagem...

Abstract:

The CORAIL project aims at defining components for Extended Modular Avionics. The contribution of project-team TEA is to define a specification method and to provide a generator of multi-task applications.

9.2. International Initiatives

9.2.1. International Project Grants

9.2.1.1. US Air Force Office for Scientific Research – Grant FA8655-13-1-3049

Title: Co-Modeling of Safety-Critical Multi-threaded Embedded Software for Multi-Core Embedded Platforms

Inria principal investigator: Jean-Pierre Talpin

International Partner (Institution - Laboratory - Researcher):

Virginia Tech Research Laboratories, Arlington (United States)

Embedded Systems Group, Technische Universität Kaiserslautern (Germany)

Duration: 2013 - 2016

See also: http://www.irisa.fr/espresso/Polycore

Abstract: The aim of the USAF OSR Grant FA8655-13-1-3049 is to support collaborative research entitled "Co-Modeling of safety-critical multi-threaded embedded software for multi-core embedded platforms" between Inria project-team ESPRESSO, the VTRL Fermat Laboratory and the TUKL embedded system research group, under the program of the Polycore associate-project.

9.2.1.2. Applied Science & Technology Research Institute (ASTRI, Hong Kong)

Title: Virtual Prototyping of Embedded Software Architectures

Inria principal investigator: Jean-Pierre Talpin

International Partner: ASTRI, Hong Kong

Duration: 2015 - 2016

Abstract: the topics of our present collaboration is essentially on heterogeneous time modeling for virtual prototyping in cyber-physical systems. Our project covers a wide spectrum of area of experience developed since 2012 and comprising

- model-based design and analysis of cyber-physical systems;
- system-level virtual prototyping and validation;
- design space exploration and system synthesis;

9.2.2. Inria International Labs

9.2.2.1. SACCADES

Title: Saccades

International Partner:

LIAMA

East China Normal University

Inria project-teams Aoste and Tea

Duration: 2003 - now

The SACCADES project is a LIAMA project hosted by East China Normal University and jointly led by Vania Joloboff (Inria) and Min Zhang (ECNU). The SACCADES project aims at improving the development of reliable cyber physical systems and more generally of distributed systems combining asynchronous with synchronous aspects, with different but complementary angles:

- develop the theoretical support for Models of Computations and Communications (MoCCs) that are the fundamentals basis of the tools.
- develop software tools (a) to enable the development and verification of executable models of the application software, which may be local or distributed and (b) to define and optimize the mapping of software components over the available resources.
- develop virtual prototyping technology enabling the validation of the application software on the target hardware platform.

The ambition of SACCADES project is to develop

- Theoretical Support for Cyber Physical Systems
- Software Tools for design and validation of CPS
- Virtual Prototyping of CPS

9.2.3. Inria International Partners

9.2.3.1. POLYCORE

Title: Models of computation for embedded software design

International Partner:

Virginia Tech Research Laboratories (USA)

University of Kanpur (India)

Duration: 2002 - now

Team TEA collaborates with Sandeep Shukla (now with IIT Kanpur) and his team at Virginia Tech, since 2002 (NSF-Inria BALBOA and Polycore projects, USAF OSR grant).

To date, our fruitful and sustained collaboration has yield the creation of the ACM-IEEE MEMOCODE conference series in 2003, of the ACM-SIGDA FMGALS workshop series, and of a full-day tutorial at ACM-IEEE DATE'09 on formal methods in system design. We have jointly edited two books with Springer [00], two special issues of the IEEE Transactions on Computers and one of the IEEE Transactions on Industrial Informatics, and published more than 40 joint journal articles and conference papers. We published a joint paper at the 52nd. Digital Automation Conference in San Francisco [11].

[0] *Formal methods and models for system design*, R. Gupta, S. Shukla, J.-P. Talpin, Eds. ISBN 1-4020-8051-4. Springer, 2004.

[0] *Synthesis of embedded systems*. S. Shukla, J.-P. Talpin, Eds. ISBN 978-1-4419-6399-4. Springer, 2010

9.2.3.2. VESA

Title: Virtual Prototyping of embedded software architectures

International Partner:

Applied Science & Technology Research Institute (ASTRI, Hong Kong)

The University of Hong Kong

Duration: 2012 - now

We collaborate with John Koo, now with ASTRI, and LIAMA since 2012 through visiting grants of the Chinese Academy of Science and of the University of Rennes on the topics of heterogeneous time modeling and virtual prototyping in cyber-physical systems.

In the context of project ITF ARD159 (System-Level Virtual Prototyping of Embedded Systems), ASTRI has used Polychrony and AADL to collaboratively develop a platform for conducting the design of an hardware-in-the-loop simulation of an UR5 robot arm, from its physical model described using Matlab/Simulink and powered using an Opal-RT/RT-Lab workstation, structured around an AADL system model, and using Polychrony to orchestrate real-time simulation down to FPGA analog outputs.

9.2.3.3. TIX

Title: Time In Cybernetic Systems

International Partner:

Rajesh Gupta, UCSD

Mani Srivastava, UCLA

Start year: 2015

The first topic of our collaboration is the formal definition of cross-domains clock models in system design and the formal verification of time stabilization and synchronization protocols used in distributed systems (sensor networks, data-bases). In this prospect, the NSF project Roseline is our basis of investigation (https://sites.google.com/site/roselineproject). Roseline aims at enabling robust, secure and efficient knowledge of time across the system stack.

Our second topic of collaboration is the refoundation of time modeling in high-level reactive and scripting languages, for application to the above using uni-kernels to cut through system stacks. We aim at applying the concepts of refinement types to formally specify and infer timing properties in CPS models from different system design view-point (physical, hardware, software) and using different levels of abstraction into multi-sorted 1st order logic (delta-decidability, linear arithmetic, Boolean logic, temporal logic).

9.3. International Research Visitors

9.3.1. Visits of International Scientists

Rajesh Gupta (UC San Diego) visited project TEA in July 2016 in the context of IIP TIX.

Brian Larson (FDA) visited project TEA in January and July 2016.

9.3.1.1. Internships

Daian Yue that was selected in the joint program between ENS Rennes and ECNU and joined project TEA for a six month internship in 2016.

9.3.2. Visits to International Teams

Vania Joloboff was invited for two short stays at University of East China Normal University in Shanghai and UC San Diego.

Jean-Pierre Talpin visited ASTRI in May and December, in the context of IIP VESA.

Jean-Pierre Talpin visited UC San Diego in October, in the context of IIP TIX.

Jean-Pierre Talpin visited IIT Kanpur in February and November for the preparation and Chair of MEMOCODE'16.

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific events organisation

10.1.1.1. General chair, scientific chair

Jean-Pierre Talpin served as General Chair and Finance Chair of the 14th. ACM-IEEE Conference on Methods and Models for System Design (MEMOCODE'16, IIT Kanpur, October 18-20.).

10.1.1.2. Member of the organizing committees

Jean-Pierre Talpin and Vania Joloboff co-organized the Shonan workshop on "Architecture-Centric Modeling, Analysis, and Verification of Cyber-Physical Systems" in collaboration with Toyota ITC and Denso, March 21-24.

Jean-Pierre Talpin is a member of the steering committee of the ACM-IEEE Conference on Methods and Models for System Design (MEMOCODE).

10.1.2. Scientific events selection

10.1.2.1. Member of the conference program committees

Jean-Pierre Talpin served the program committee of:

- ACVI'16, 3rd. Workshop on Architecture Centric Virtual Integration
- HLDVT'16, 18th. IEEE International High-Level Design Validation and Test Workshop
- ICESS'16, 13th. IEEE International Conference on Embedded Software and Systems
- IDEA'16, 2nd. International Workshop Integrating Data-flow, Embedded computing and Architecture
- LCTES'16, 19th. ACM SIGPLAN-SIGBED Conference on Languages, Compilers, and Tools for Embedded Systems
- MEMOCODE'16, 14th. ACM-IEEE Conference on Methods and Models for System Design
- SAC'16, 31st. ACM SIGAPP Symposium on Applied Computing
- SCOPES'16, 19th. International Workshop on Software and Compilers for Embedded Systems
- TASE'16, 10th. Theoretical Aspects of Software Engineering Conference

10.1.3. Journal

10.1.3.1. Member of the editorial boards

Jean-Pierre Talpin is Associate Editor with the ACM Transactions for Embedded Computing Systems (TECS), with the Springer journal on Frontiers of Computer Science (FCS) and with the EURASIP journal of embedded systems (JES).

10.1.3.2. reviewer

Jean-Pierre Talpin reviewed articles for Acta Informatica.

Thierry Gautier reviewed for Frontiers of Computer Science.

10.2. Teaching - Supervision - Juries

10.2.1. Invited talks

Vania Joloboff gave a talk in the series of the Distinguished Lecturers of the Computer Science and Engineering department at UC San Diego.

Jean-Pierre Talpin gave an invited presentation at the APAC 2016 Summit on Robotics at the HKSTP in Hong Kong https://www.apacinnosummit.net.

10.2.2. Supervision

Vania Joloboff supervised work of Master student Daian Yue that was selected in the joint program between ENS Rennes and ECNU.

Jean-Pierre Talpin is the supervisor of Simon Lunel's thesis on "*Timed contract algebras for correct by construction real-time system design*".

10.2.3. Juries

Jean-Pierre Talpin served as examiner for the Ph.D. Thesis defense of Fatma Jebali on "Formal Framework for modeling and Verifying Globally Asynchronous Locally Synchronous Systems", September 12., in Grenoble.

Jean-Pierre Talpin served as referee for the PhD. Thesis Defence of Amani Khecharem on "Une approche de méta-modélisation pour la représentation multi-vues des architectures hétérogènes embarqués", May 3., in Sophia Antipolis.

11. Bibliography

Major publications by the team in recent years

[1] L. BESNARD, T. GAUTIER, P. LE GUERNIC, J.-P. TALPIN.*Compilation of Polychronous Data Flow Equations*, in "Synthesis of Embedded Software", S. K. SHUKLA, J.-P. TALPIN (editors), Springer, 2010, p. 1-40 [*DOI :* 10.1007/978-1-4419-6400-7_1], http://hal.inria.fr/inria-00540493.

[2] A. BOUAKAZ, J.-P. TALPIN.*Design of Safety-Critical Java Level 1 Applications Using Affine Abstract Clocks*, in "International Workshop on Software and Compilers for Embedded Systems", St. Goar, Germany, June 2013, p. 58-67 [*DOI :* 10.1145/2463596.2463600], https://hal.inria.fr/hal-00916487.

[3] C. BRUNETTE, J.-P. TALPIN, A. GAMATIÉ, T. GAUTIER.*A metamodel for the design of polychronous systems*, in "The Journal of Logic and Algebraic Programming", 2009, vol. 78, n° 4, p. 233 - 259, IFIP WG1.8 Workshop on Applying Concurrency Research in Industry [*DOI :* 10.1016/J.JLAP.2008.11.005], http://www.sciencedirect.com/science/article/pii/S1567832608000957.

[4] A. GAMATIÉ, T. GAUTIER, P. LE GUERNIC, J.-P. TALPIN.*Polychronous Design of Embedded Real-Time Applications*, in "ACM Transactions on Software Engineering and Methodology (TOSEM)", April 2007, vol. 16, n° 2, http://doi.acm.org/10.1145/1217295.1217298.

[5] A. GAMATIÉ, T. GAUTIER.*The Signal Synchronous Multiclock Approach to the Design of Distributed Embedded Systems*, in "IEEE Transactions on Parallel and Distributed Systems", 2010, vol. 21, n° 5, p. 641-657 [*DOI :* 10.1109/TPDS.2009.125], http://hal.inria.fr/inria-00522794.

[6] A. GAMATIÉ, T. GAUTIER, P. LE GUERNIC.*Synchronous design of avionic applications based on model refinements*, in "Journal of Embedded Computing (IOS Press)", 2006, vol. 2, n° 3-4, p. 273-289, http://hal.archives-ouvertes.fr/hal-00541523.

[7] T. GAUTIER, C. GUY, A. HONORAT, P. LE GUERNIC, J.-P. TALPIN, L. BESNARD.*Polychronous Automata and their Use for Formal Validation of AADL Models*, in "Frontiers of Computer Science -Springer-", December 2016, https://hal.inria.fr/hal-01411257.

[8] P. LE GUERNIC, J.-P. TALPIN, J.-C. LE LANN.*Polychrony for system design*, in "Journal of Circuits, Systems and Computers, Special Issue on Application Specific Hardware Design", June 2003, vol. 12, n^o 03, http://hal.inria.fr/docs/00/07/18/71/PDF/RR-4715.pdf.

[9] D. POTOP-BUTUCARU, Y. SOREL, R. DE SIMONE, J.-P. TALPIN.*From Concurrent Multi-clock Programs to Deterministic Asynchronous Implementations*, in "Fundamenta Informaticae", January 2011, vol. 108, n^o 1-2, p. 91–118, http://dl.acm.org/citation.cfm?id=2362088.2362094.

[10] J.-P. TALPIN, J. OUY, T. GAUTIER, L. BESNARD, P. LE GUERNIC.*Compositional design of isochronous systems*, in "Science of Computer Programming", February 2012, vol. 77, n^o 2, p. 113-128 [*DOI :* 10.1016/J.SCICO.2010.06.006], http://hal.archives-ouvertes.fr/hal-00768341.

[11] H. YU, J. PRASHI, J.-P. TALPIN, S. K. SHUKLA, S. SHIRAISHI.*Model-Based Integration for Automotive Control Software*, in "Digital Automation Conference", San Francisco, United States, ACM, June 2015, https://hal.inria.fr/hal-01148905.

Publications of the year

Articles in International Peer-Reviewed Journal

[12] T. GAUTIER, C. GUY, A. HONORAT, P. LE GUERNIC, J.-P. TALPIN, L. BESNARD.*Polychronous Automata and their Use for Formal Validation of AADL Models*, in "Frontiers of Computer Science -Springer-", December 2016, https://hal.inria.fr/hal-01411257.

International Conferences with Proceedings

[13] D. BAELDE, S. LUNEL, S. SCHMITZ.*A Sequent Calculus for a Modal Logic on Finite Data Trees*, in "CSL 2016", Marseille, France, J.-M. TALBOT, L. REGNIER (editors), Leibniz International Proceedings in Informatics, LZI, September 2016, vol. 62, n^o 32, p. 1–16 [*DOI :* 10.4230/LIPICS.CSL.2016.32], https://hal.inria.fr/hal-01191172.

[14] L. BESNARD, T. GAUTIER, C. GUY, P. LE GUERNIC, J.-P. TALPIN, B. R. LARSON, E. BORDE.*Formal semantics of behavior specifications in the architecture analysis and design language standard*, in "18th IEEE International High-Level Design Validation and Test Workshop", Santa Cruz, United States, High-level design, verification and test, IEEE, October 2016, p. 30 - 39 [*DOI :* 10.1109/HLDVT.2016.7748252], https://hal.inria.fr/hal-01419968.

[15] C. JUNKE, T. GAUTIER, J.-P. TALPIN, L. BESNARD.*Integration of polychrony and QGen model compiler*, in "ERTS'16 - European Congress on Embeddd Real-Rime Software and Systems", Toulouse, France, January 2016, https://hal.inria.fr/hal-01241808.

[16] V. C. NGO, A. LEGAY, V. JOLOBOFF.*PSCV: A Runtime Verification Tool for Probabilistic SystemC Models*, in "CAV 2016 - 28th International Conference on Computer Aided Verification", Toronto, Canada, S. CHAUDHURI, A. FARZAN (editors), LNCS - Lecture Notes in Computer Science, Springer, July 2016, vol. 9779, p. 84 - 91 [*DOI :* 10.1007/978-3-319-41528-4_5], https://hal.inria.fr/hal-01406488.

[17] D. YUE, V. JOLOBOFF, F. MALLET.*Flexible Runtime Verification Based On Logical Clock Constraints*, in "FDL 2016 - Forum on specification & Design Languages", Bremen, Germany, September 2016, https://hal.inria.fr/hal-01421890.

Conferences without Proceedings

[18] O. AUMAGE, D. BARTHOU, A. HONORAT.*A Stencil DSEL for Single Code Accelerated Computing with SYCL*, in "SYCL 2016 1st SYCL Programming Workshop during the 21st ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming", Barcelone, Spain, March 2016, https://hal.archives-ouvertes.fr/hal-01290099.

[19] O. SANKUR, J.-P. TALPIN.*An Abstraction Technique For Parameterized Model Checking of Leader Election Protocols: Application to FTSP*, in "23rd International Conference on Tools and Algorithms for the Construction and Analysis of Systems (TACAS)", Uppsala, Sweden, April 2017, https://hal.archives-ouvertes.fr/hal-01431472.

Research Reports

[20] L. BESNARD, T. GAUTIER, P. LE GUERNIC, C. GUY, J.-P. TALPIN, B. LARSON, E. BORDE.*Formal semantics of behavior specifications in the architecture analysis and design language standard*, Inria, 2016, n^{o} 8950, p. 30 - 39 [*DOI :* 10.1109/HLDVT.2016.7748252], https://hal.inria.fr/hal-01419973.

Project-Team VISAGES

Vision, Action and information manaGement System in health

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

IN PARTNERSHIP WITH:
CNRS

INSERM

Université Rennes 1

RESEARCH CENTER
Rennes - Bretagne-Atlantique

THEME
Computational Neuroscience and Medecine

Table of contents

Project-Team VISAGES

Creation of the Project-Team: 2005 July 04

Keywords:

Computer Science and Digital Science:

3.1.2. - Data management, quering and storage
3.1.3. - Distributed data
3.1.7. - Open data
3.1.8. - Big data (production, storage, transfer)
3.2.4. - Semantic Web
3.3.3. - Big data analysis
3.4.1. - Supervised learning
3.4.2. - Unsupervised learning
3.4.3. - Reinforcement learning
3.4.4. - Optimization and learning
3.4.7. - Kernel methods
5.1.4. - Brain-computer interfaces, physiological computing
5.2. - Data visualization
5.3.2. - Sparse modeling and image representation
5.3.3. - Pattern recognition
5.3.4. - Registration
5.4.1. - Object recognition
5.4.5. - Object tracking and motion analysis
5.4.6. - Object localization
5.9.2. - Estimation, modeling
6.2.3. - Probabilistic methods
6.2.4. - Statistical methods
6.3.3. - Data processing

Other Research Topics and Application Domains:

1.3. - Neuroscience and cognitive science
1.3.1. - Understanding and simulation of the brain and the nervous system
1.3.2. - Cognitive science
1.4. - Pathologies
2.1. - Well being
2.2.6. - Neurodegenerative diseases
2.5.1. - Sensorimotor disabilities
2.5.2. - Cognitive disabilities
2.6.1. - Brain imaging

1. Members

Research Scientists

Christian Barillot [Team leader, CNRS, Senior Researcher, HDR]
Emmanuel Caruyer [CNRS, Researcher]
Olivier Commowick [Inria, Researcher]
Sylvain Prima [Inria, Researcher]

Faculty Members

Isabelle Bonan [Univ. Rennes I, Professor, HDR]
Gilles Edan [Univ. Rennes I, Professor, HDR]
Jean-Christophe Ferré [Univ. Rennes I, Professor, HDR]
Jean-Yves Gauvrit [Univ. Rennes I, Professor, HDR]
Pierre Maurel [Univ. Rennes I, Assistant Professor]

Technical Staff

Élise Bannier [CHRU Rennes]
Isabelle Corouge [Univ. Rennes I]
Yao Chi [Inria, granted by ANR France Life Imaging project]
Rene-Paul Debroize [Inria, until Jul 2016, granted by MEDDAY PHARMACEUTICALS]
Inès Fakhfakh [Inria, granted by OFSEP Univ. Claude Bernard Lyon 1]
Justine Guillaumont [Inria, until Jan 2016, granted by OFSEP Univ. Claude Bernard Lyon 1]
Michael Kain [Inria, granted by ANR France Life Imaging project]
Florent Leray [Inria, granted by ANR France Life Imaging project]
Julien Louis [Inria, since Apr 2016, granted by ANR France Life Imaging project]
Marsel Mano [Inria, 50% with Hybrid project-team, granted by ANR CominLabs "Hemisfer"]
Aneta Morawin [Inria, since Mar 2016, granted by ANR France Life Imaging project]
Mathieu Simon [Inria, granted by ANR France Life Imaging project]
Benoit Combès [Inria, granted by EMISEP project]

PhD Students

Sudhanya Chatterjee [Univ. Rennes I]
Hrishikesh Deshpande [Inria, until Jul 2016]
Renaud Hédouin [Inria, granted by Conseil Régional de Bretagne]
Pierre-Yves Jonin [CHRU Rennes]
Anne Kerbrat [CHRU Rennes]
Antoine Legouhy [CNRS, since Nov 2016]
Cédric Meurée [Siemens Healthineers, granted by CIFRE Inria / Siemens]
Lorraine Perronnet [Inria, granted by Conseil Régional de Bretagne]
Maia Proisy [CHRU Rennes]
Haykel Snoussi [Inria, granted by Conseil Régional de Bretagne and EMISEP project]
Corentin Vallée [Univ. Rennes I, since Nov 2016]

Post-Doctoral Fellows

Jean-Marie Batail [CHRU Rennes, MD]
Quentin Duché [Inria, granted by INCR foundation]
Saman Noorzadeh [Inria, granted by Labex CominLabs]

Visiting Scientists

Michael Paquette [Univ de Sherbrooke, from Jun to Aug 2016, granted by MITACS-Inria]
Lorenzo Rota [since Oct 2016]

Administrative Assistant

Angélique Jarnoux [Inria]

Others

Hélène Raoult [CHRU Rennes, until Jun 2016]
Florence Prigent Le Jeune [CRLCC Rennes, since Jul 2016]

2. Overall Objectives

2.1. Overall objectives

Medical Imaging, Neuroinformatics, Neuroimaging, Medical Image Computing, Modeling of normal and pathological behavior of the human brain, e-health & HealthGrids

The Unit/Project VISAGES U746 is a research team jointly affiliated to INSERM (National Institute of Health and Scientific Research), Inria (National Institute of Research in Computer Sciences and Automation) and IRISA / UMR CNRS 6074, University of Rennes I. We are located in Rennes, France on both medical and sciences campus. The team has been created in 2005. Our ambition is to set up a multidisciplinary team merging researchers in image processing and medical doctors. The goal of VISAGES is to constitute a multidisciplinary team. Even though, research in medical imaging could find motivation and recognition based on methodological breakthroughs alone, the ultimate goal, when dealing with medical imaging research, is to make the clinical practice benefit from the basic and applied research, while keeping the excellence of the methodological research. This objective entails the creation of teams encompassing clinical and scientific researchers to design and conduct research projects together. Our aim through the past period was to build a research team able to perform a research going from a novel and basic stage to original clinical experimentation with clear medical impact.

Our research activities are focused on the research and development of new algorithms in medical imaging in the context of the pathologies of the central nervous system. In this context, we are addressing the general problems of the better understanding of normal and pathological brain organs and systems behavior, at different scales, and the promotion and the support of Virtual Organizations of biomedical actors by means of healthgrid's technologies. The medical application objectives are focused on pathologies of the central nervous system, with a particular effort on extraction of new imaging biomarkers for brain pathologies (e.g. Multiple Sclerosis, neuropaediatrics, strokes, psychiatry, ...). More generally, our application objectives concern the following diseases: Multiple sclerosis, epilepsy, dementia, neuro-degenerative brain diseases, brain vascular diseases.

3. Research Program

3.1. Research Program

The scientific foundations of our team concern the development of new processing algorithms in the field of medical image computing : image fusion (registration and visualization), image segmentation and analysis, management of image related information. Since this is a very large domain, which can endorse numerous types of application; for seek of efficiency, the purpose of our methodological work primarily focuses on clinical aspects and for the most part on head and neck related diseases. In addition, we emphasize our research efforts on the neuroimaging domain. Concerning the scientific foundations, we have pushed our research efforts:

- In the field of image fusion and image registration (rigid and deformable transformations) with a special emphasis on new challenging registration issues, especially when statistical approaches based on joint histogram cannot be used or when the registration stage has to cope with loss or appearance of material (like in surgery or in tumor imaging for instance).
- In the field of image analysis and statistical modeling with a new focus on image feature and group analysis problems. A special attention was also to develop advanced frameworks for the construction of atlases and for automatic and supervised labeling of brain structures.
- In the field of image segmentation and structure recognition, with a special emphasis on the difficult problems of *i*) image restoration for new imaging sequences (new Magnetic Resonance Imaging protocols, 3D ultrasound sequences...), and *ii*) structure segmentation and labelling based on shape, multimodal and statistical information.
- Following the Neurobase national project where we had a leading role, we wanted to enhance the development of distributed and heterogeneous medical image processing systems.

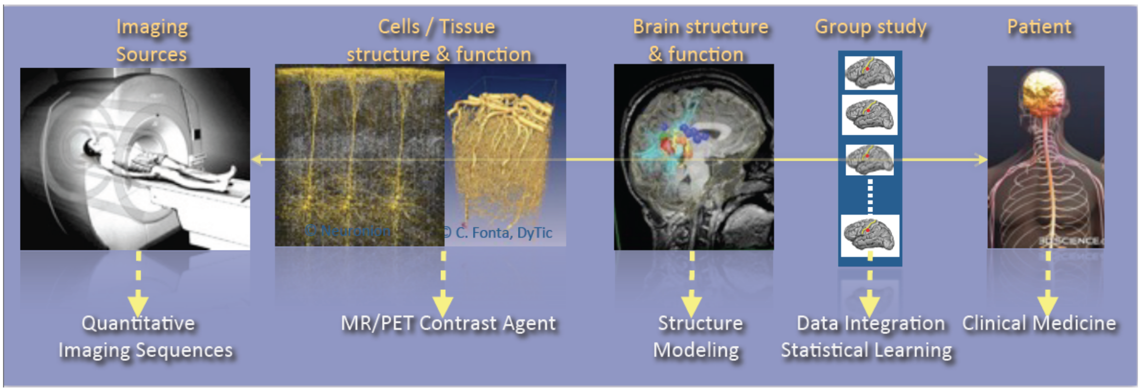


Figure 1. The major overall scientific foundation of the team concerns the integration of data from the Imaging source to the patient at different scales: from the cellular or molecular level describing the structure and function, to the functional and structural level of brain structures and regions, to the population level for the modelling of group patterns and the learning of group or individual imaging markers.

As shown in Fig. 1, research activities of the VISAGES U746 team are tightly coupling observations and models through integration of clinical and multi-scale data, phenotypes (cellular, molecular or structural patterns). We work on personalized models of central nervous system organs and pathologies, and intend to confront these models to clinical investigation studies for quantitative diagnosis, prevention of diseases, therapy planning and validation. These approaches are developed in a translational framework where the data integration process to build the models inherits from specific clinical studies, and where the models are assessed on prospective clinical trials for diagnosis and therapy planning. All of this research activity is conducted in tight links with the Neurinfo imaging platform environments and the engineering staff of the platform. In this context, some of our major challenges in this domain concern:

- The elaboration of new descriptors to study the brain structure and function (e.g. variation of brain perfusion with and without contrast agent, evolution in shape and size of an anatomical structure in relation with normal, pathological or functional patterns, computation of asymmetries from shapes and volumes).
- The integration of additional spatio-temporal imaging sequences covering a larger range of observation, from the molecular level to the organ through the cell (Arterial Spin Labeling, diffusion MRI, MR relaxometry, MR cell labeling imaging, PET molecular imaging, ...). This includes the elaboration of new image descriptors coming from spatio-temporal quantitative or contrast-enhanced MRI.
- The creation of computational models through data fusion of molecular, cellular, structural and functional image descriptors from group studies of normal and/or pathological subjects.
- The evaluation of these models on acute pathologies especially for the study of degenerative, psychiatric or developmental brain diseases (e.g. Multiple Sclerosis, Epilepsy, Parkinson, Dementia, Strokes, Depression, Schizophrenia, ...) in a translational framework.

In terms of methodological developments, we are particularly working on statistical methods for multidimensional image analysis, and feature selection and discovery, which includes:

- The development of specific shape and appearance models, construction of atlases better adapted to a patient or a group of patients in order to better characterize the pathology;
- The development of advanced segmentation and modeling methods dealing with longitudinal and

multidimensional data (vector or tensor fields), especially with the integration of new prior models to control the integration of multiscale data and aggregation of models;

- The development of new models and probabilistic methods to create water diffusion maps from MRI;
- The integration of machine learning procedures for classification and labeling of multidimensional features (from scalar to tensor fields and/or geometric features): pattern and rule inference and knowledge extraction are key techniques to help in the elaboration of knowledge in the complex domains we address;
- The development of new dimensionality reduction techniques for problems with massive data, which includes dictionary learning for sparse model discovery. Efficient techniques have still to be developed to properly extract from a raw mass of images derived data that are easier to analyze.

4. Application Domains

4.1. Neuroimaging

One research objective in neuroimaging is the construction of anatomical and functional cerebral maps under normal and pathological conditions. Many researches are currently performed to find correlations between anatomical structures, essentially sulci and gyri, where neuronal activation takes place, and cerebral functions, as assessed by recordings obtained by the means of various neuroimaging modalities, such as PET (Positron Emission Tomography), fMRI (Functional Magnetic Resonance Imaging), EEG (Electro-EncephaloGraphy) and MEG (Magneto-EncephaloGraphy). Then, a central problem inherent to the formation of such maps is to put together recordings obtained from different modalities and from different subjects. This mapping can be greatly facilitated by the use of MR anatomical brain scans with high spatial resolution that allows a proper visualization of fine anatomical structures (sulci and gyri). Recent improvements in image processing techniques, such as segmentation, registration, delineation of the cortical ribbon, modeling of anatomical structures and multi-modality fusion, make possible this ambitious goal in neuroimaging. This problem is very rich in terms of applications since both clinical and neuroscience applications share similar problems. Since this domain is very generic by nature, our major contributions are directed towards clinical needs even though our work can address some specific aspects related to the neuroscience domain.

4.2. Multiple sclerosis

Over the past years, a discrepancy became apparent between clinical Multiple sclerosis (MS) classification describing on the one hand MS according to four different disease courses and, on the other hand, the description of two different disease stages (an early inflammatory and a subsequently neurodegenerative phase). It is to be expected that neuroimaging will play a critical role to define in vivo those four different MS lesion patterns. An in vivo distinction between the four MS lesion patterns, and also between early and late stages of MS will have an important impact in the future for a better understanding of the natural history of MS and even more for the appropriate selection and monitoring of drug treatment in MS patients. MRI has a low specificity for defining in more detail the pathological changes which could discriminate between the different lesion types. However, it has a high sensitivity to detect focal and also widespread, diffuse pathology of the normal appearing white and gray matter. Our major objective within this application domain is then to define new neuroimaging markers for tracking the evolution of the pathology from high dimensional data (e.g. nD+t MRI) in the brain and the spinal cord. In addition, in order to complement MR neuroimaging data, we ambition to perform also cell labeling neuroimaging (e.g. MRI or PET) and to compare MR and PET data using standard and experimental MR contrast agents and radiolabeled PET tracers for activated microglia (e.g. USPIO or PK 11195). The goal is to define and develop, for routine purposes, cell specific and also quantitative imaging markers for the improved in vivo characterization of MS pathology.

4.3. Modeling of anatomical and anatomo-functional neurological patterns

The major objective within this application domain is to build anatomical and functional brain atlases in the context of functional mapping and for the study of developmental, neurodegenerative or even psychiatric brain diseases (Multiple sclerosis, Epilepsy, Parkinson, Dysphasia, Depression or even Alzheimer). This is a very competitive research domain; our contribution is based on our previous works in this field, and by continuing our local and wider collaborations.

An additional objective within this application domain is to find new descriptors to study the brain anatomy and/or function (e.g. variation of brain perfusion, evolution in shape and size of an anatomical structure in relation with pathology or functional patterns, computation of asymmetries ...). This is also a very critical research domain, especially for many developmental or neurodegenerative brain diseases.

5. Highlights of the Year

5.1. Highlights of the Year

5.1.1. Awards

- In 2015, the Neurinfo platform obtained an “Emergence” label from the IBISA agency, this label has been upgraded in 2016 as a “platform of Excellence” and sustained by IBISA in 2016 and onward. The IBISA label is a national label for technological platforms awarded by the GIS IBISA on an annual basis.

6. New Software and Platforms

6.1. Anima

KEYWORDS: Filtering - Medical imaging - Diffusion imaging - Registration - Relaxometry
SCIENTIFIC DESCRIPTION: Anima is a set of libraries and tools developed by the team as a common repository of research algorithms. As of now, it contains tools for image registration, statistical analysis (group comparison, patient to group comparison), diffusion imaging (model estimation, tractography, etc.), quantitative MRI processing (quantitative relaxation times estimation, MR simulation), image denoising and filtering, and segmentation tools. All of these tools are based on stable libraries (ITK, VTK), making it simple to maintain.

- Participants: Olivier Commowick, Rene-Paul Debroize, Florent Leray and Renaud Hédouin
- Contact: Olivier Commowick
- APP number: IDDN.FR.001.460020.000.S.P.2015.000.31230
- URL: https://github.com/Inria-Visages/Anima-Public/wiki

6.2. MedInria

KEYWORDS: Segmentation - Health - DWI - Visualization - Medical imaging
SCIENTIFIC DESCRIPTION: It aims at creating an easily extensible platform for the distribution of research algorithms developed at Inria for medical image processing. This project has been funded by the D2T (ADT MedInria-NT) in 2010 and renewed in 2012. The VisAGeS team leads this Inria national project and participates in the development of the common core architecture and features of the software as well as in the development of specific plugins for the team’s algorithm.

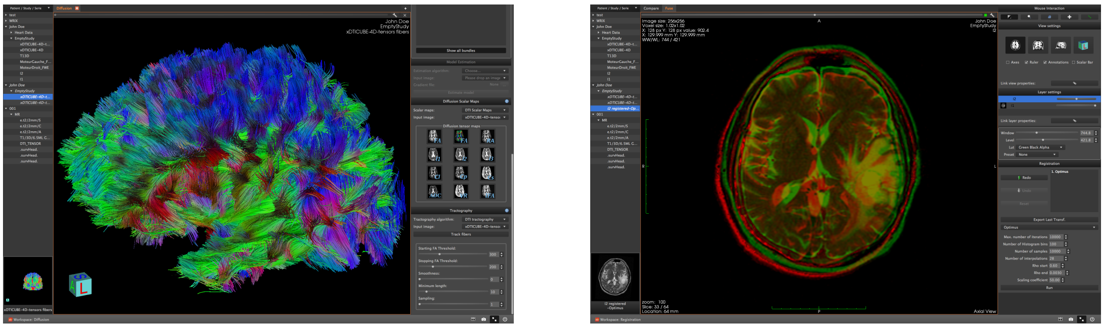


Figure 2. The medInria software platform: Tractography overlapped with 3D image (left), and Fused view of registered images (right).

FUNCTIONAL DESCRIPTION: MedInria is a free software platform dedicated to medical data visualization and processing as illustrated in Fig. 2.

- Participants: Olivier Commowick and Rene-Paul Debroize
- Partners: HARVARD Medical School - IHU - LIRYC - IHU - Strasbourg - NIH
- Inria structures involved : ASCLEPIOS, ATHENA, PARIETAL, VISAGES
- Contact: Olivier Commowick
- URL: http://med.inria.fr
- APP number: IDDN.FR.001.130017.000.S.A.2012.000.31230

6.3. Shanoir

SHAring NeurOImaging Resources
KEYWORDS: Shanoir - Webservices - Data base - Biology - Health - DICOM - Neuroimaging - Medical imaging - PACS - Nifti - Data Sharing - Web Application
FUNCTIONAL DESCRIPTION: SHAring NeurOImaging Resources (Shanoir, Previously InriaNeuroTk) is an open source software platform designed to structure, manage, archive, visualize and share neuroimaging data with an emphasis on multi-centric collaborative research projects. It provides common features of neuroimaging data management systems along with research-oriented data organization and enhanced accessibility (see Fig. 3).

Shanoir is a secured J2EE application running on a JBoss server, reachable via graphical interfaces in a browser or by third party programs via web services. It behaves as a repository of neuroimaging files coupled with a relational database hoding meta-data. The data model, based on OntoNeurolog, an ontology devoted to the neuroimaging field, is structured around research studies where of involved patients have examinations which either produce image acquisitions or clinical scores. Each image acquisition is composed of datasets represented by their acquisition parameters and image files. The system only keeps anonymous data.

Image files imports are possible from various sources (DICOM CDs, PACs, image files in NIfTI / Analyze format) using either online wizards, with completions of related meta-data, or commande line tools. Once de-identified during the import phase, DICOM header's customizable feature. Shanoir can also record any executed processing allowing to retrieve workflows applied to a particular dataset along with the intermediate data.

The clinical scores resulting from instrument based assessments (e.g. neuropsychological tests) can also be entered and easily retrieved and exported in different formats (Excel, CSV, Xml). Scores and image acquisitions are bound together which makes relationship analysis possible. The instrument database is scalable an new measures can be added in order to meet specific project needs, by use of intuitive graphical interfaces.

Using cross-data navigation and advanced search criteria, the users can quickly point to a subset of data to be downloaded. Client side applications have as well been developed to illustrate how to locally access and exploit data through the available web services. With regards to security, the system requires authentication and user rights are tunable for each hosted studies. A study responsible can thereby define the users allowed to see, download or import data into his study or simply make it public.

Shanoir serves neuroimaging researchers in organizing efficiently their studies while cooperating with other laboratories. By managing patient privacy, Shanoir allows the exploitation of clinical data in a research context. It is finally a handy solution to publish and share data with a broader community.

Shanoir integrates the enterprise search platform, Apache Solr, to provide the users a vast array of advanced features such as near real-time indexing and queries, full-text search, faceted navigation, autosuggestion and autocomplete.

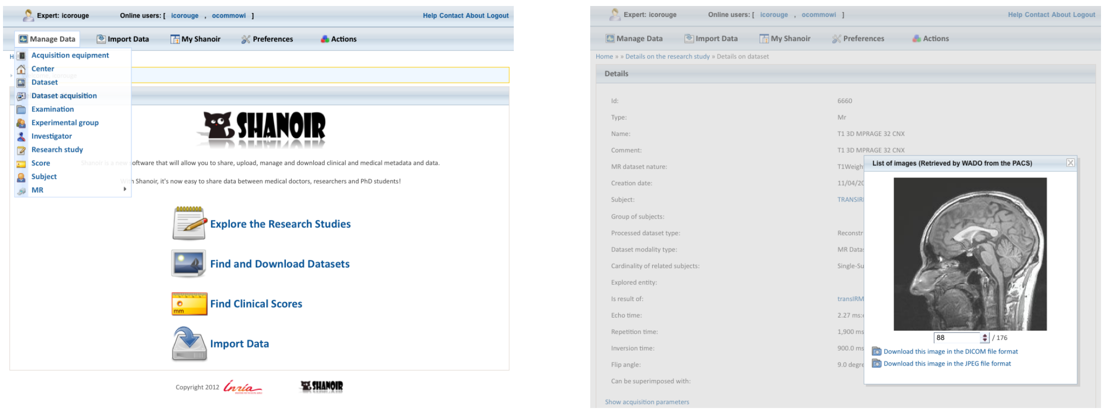


Figure 3. The SHANOIR software is a web application to share, archive, search and visualize neuroimaging data.

- Participants: Michael Kain, Christian Barillot, Anthony Baire, Mathieu Simon, Julien Louis, Isabelle Corouge, Élise Bannier, Aneta Morawin and Yao Chi
- Partners: CNRS - INSERM - Université de Rennes 1
- Contact: Christian Barillot
- URL: http://shanoir.gforge.inria.fr
- APP number: IDDN.FR.001.520021.003.S.A.2008.000.31230 (2014/08/20)

6.4. QtShanoir

KEYWORDS: Shanoir - Qt - Webservices - Soap - C++ - Health - DICOM - Plug-in - Medical imaging - Nifti

SCIENTIFIC DESCRIPTION: QtShanoir is based on Qt/C++ librairie. It interacts with the Shanoir server using SOAP web services provided. This application queries the server and displays hierarchical data extracted in tree view. Data could also be easily downloaded or uploaded on the server. In order to extend the Shanoir environment, QtShanoir is developed to contain two shared libraries:

- "GUI" that represents all user interfaces.
- "DAO" that takes in charge the data model. This library assures the connection to the server and provides all QtShanoir services : research, download and upload of Processed Dataset (NIfTI).

QtShanoir dynamic libraries are already reused and integrated in other projects: in the software medInria and in an under development command line program.

FUNCTIONAL DESCRIPTION: QtShanoir is a graphical client application of the medical imaging database Shanoir. This application provides various functionalities to satisfy researchers' needs. It allows users to:

- explore neuroimaging data derived from multicenter research trials. Through an intuitive user interface, users could easily visualize voluminous amount of structured data: studies, patients and datasets extracted from Shanoir
- download and to upload data from the server.

This application is available on Windows, UNIX, MacOs X. It is integrated as a plugin in medInria, a multi-plateform for medical image processing and visualization.

- Participants: Olivier Commowick and Florent Leray
- Contact: Christian Barillot
- URL: http://qtshanoir.gforge.inria.fr
- APP number: IDDN.FR.001.130017.000.S.A.2012.000.31230 (2012/02/08)

6.5. ShanoirUploader

KEYWORDS: Shanoir - Webservices - Java - Biology - Health - DICOM - Neuroimaging - Medical imaging - PACS

SCIENTIFIC DESCRIPTION: ShanoirUploader is a desktop application on base of JavaWebStart (JWS). The application can be downloaded and installed using an internet browser. It interacts with a PACS to query and retrieve the data stored on it as illustrated in Fig. 4. After this ShanoirUploader sends the data to a Shanoir server instance in order to import these data. This application bypasses the situation, that in most of the clinical network infrastructures a server to server connection is complicated to set up between the PACS and a Shanoir server instance.

FUNCTIONAL DESCRIPTION: ShanoirUploader is a Java desktop application that transfers data securely between a PACS and a Shanoir server instance (e.g., within a hospital). It uses either a DICOM query/retrieve connection or a local CD/DVD access to search and access images from a local PACS or the local CD/DVD. After having retrieved the data, the DICOM files are locally anonymized and then uploaded to the Shanoir server. A possible integration of a hash creation application for patient identifiers is provided as well. The primary goals of that application are to enable mass data transfers between different remote server instances and therefore reduce the waiting time of the users, when importing data into Shanoir. Most of the time during import is spent with data transfers.

- Participants: Michael Kain, Inès Fakhfakh and Christian Barillot
- Contact: Christian Barillot
- URL: http://shanoir.gforge.inria.fr
- APP number: IDDN.FR.001.380026.000.S.P.2015.000.31230 (2015/09/11)

6.6. iShanoir

KEYWORDS: Shanoir - Biology - Health - Neuroimaging - Mobile application - Medical imaging - Biomedical imaging

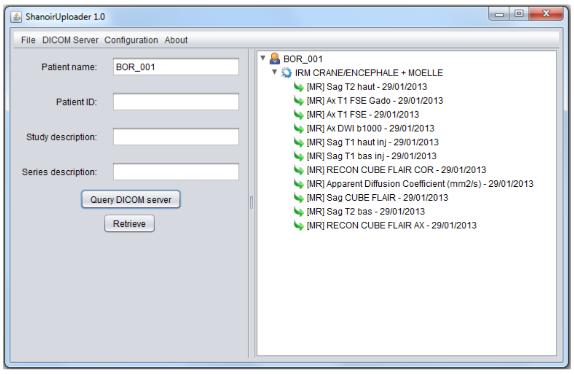


Figure 4. The ShanoirUploader software is a desktop application designed to interact with a PACS to query and retrieve the data stored on any PACS.

FUNCTIONAL DESCRIPTION iShanoir is an iOS application, designed for iPhone and iPad. On base of this application a Shanoir server can be accessed as illustrated in Fig. 5. For this the Shanoir SOAP web-services are called. iShanoir can be used to access and navigate in the data tree structure, stored on a Shanoir server. iShanoir displays as well additional meta data corresponding to the data entities in the tree structure. On base of these informations image files (NIfTI and DICOM) can be selected and downloaded on a local iPhone/iPad in a temporary cache. From this cache the files can be opened and displayed with a corresponding viewer, the user already has to have installed on his device.

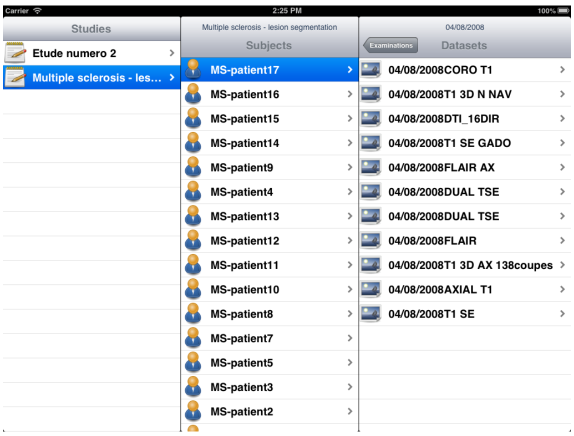


Figure 5. The iShanoir interface, showing the browsing tabs within the research studies stored in the Shanoir database.

- Participants: Michael Kain and Christian Barillot
- Contact: Christian Barillot
- URL: http://shanoir.gforge.inria.fr

6.7. autoMRI

KEYWORDS: FMRI - MRI - ASL - FASL - SPM - Automation

SCIENTIFIC DESCRIPTION: Automri is an analysis pipeline to process morphological, perfusion, BOLD fMRI, relaxometry and neurovascular data. This software is highly configurable in order to fit to a wide range of needs. Pre-processing includes segmentation of anatomical data, as well as co-registration, spatial normalization and atlas building of all data types. The analysis pipelines perform either within-group analysis or between-group or one subject-versus-group comparison and produce statistical maps of regions with significant differences. These pipelines can be applied to structural data to exhibit patterns of atrophy or lesions, to ASL (both pulsed or pseudo-continuous sequences) or PET data to detect perfusion or metabolic abnormalities, to relaxometry data to detect deviations from a template, to functional data - either BOLD or ASL - to outline brain activations related to block or event-related paradigms. In addition to the standard General Linear Model approach, the ASL pipelines implement an a contrario approach and, for patient-specific perfusion study, an heteroscedastic variance model. Besides, the vascular pipeline processes 4D MRA data and enables accurate assessment of hemodynamic patterns.

FUNCTIONAL DESCRIPTION: Based on MATLAB and the SPM8 toolbox, autoMRI provides complete pipelines to pre-process and analyze various types of images (anatomical, functional, perfusion, metabolic, relaxometry, vascular).

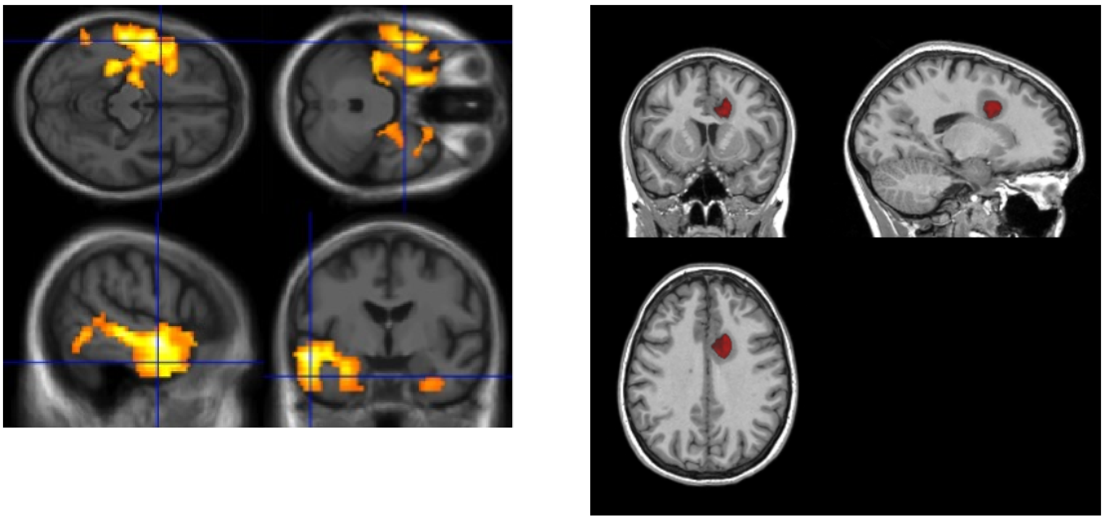


Figure 6. Illustrations of results obtained with autoMRI: Conjunction map showing areas of hypoperfusion and hypometabolism in semantic dementia (left) and detection of relaxometry defect in an MS patient (right).

- Participants: Isabelle Corouge, Quentin Duché, Cédric Meurée, Pierre Maurel and Élise Bannier.
- Contact: Isabelle Corouge
- URL: http://www.irisa.fr/visages/
- APP number: Part in IDDN.FR.001.130017.000.S.A.2012.000.31230

6.8. Integration of EEG and fMRI

KEYWORDS: medical imaging - EEG - fMRI

FUNCTIONAL DESCRIPTION: Related to the project Hemisfer there have been development of new functions, scripts and demos for the acquisition and processing of the EEG and fMRI data in Real-time. These include:

- Functions for fMRI header info reader, volume reader, motion correction, slice time correction nifty output conversion, real time fMRI initialization, real time fMRI processing, z-score calculation, volume smoother, alignment, etc., functions for real time EEG data acquisition, filtering, power calculation and display.
- Scripts for various protocols used in offline fMRI experiments, real time processing loop for EEG and fMRI.
- Demo for real time acquisition of the EEG and fMRI data, demo for real time processing efficiency of the fMRI data, demo for the real time processing of EEG data, real time z-Score for fMRI data.
- Several small aux functions for I/O interfaces (e.g. com, serial)

In the current stage the prototype also relies on various other free toolboxes (e.g. SPM, pnet).

- Participants: Marsel Mano, Lorraine Perronnet, Anatole Lecyuer, Christian Barillot.
- Contact: Marsel Mano
- International Patent Number: PCT/EP2016/1652279

6.9. Platforms

6.9.1. The Neurinfo Platform

VisAGeS is the founding actor of an experimental research platform which was installed in August 2009 at the University Hospital of Rennes. The University of Rennes 1, Inria, Inserm for the academic side, and the University Hospital of Rennes and the Cancer Institute "Eugene Marquis" for the clinical side, are partners of this neuroinformatics platform called Neurinfo [0]. This platform has been supported under the "Contrat de Projets Etat-Région" (Christian Barillot is the PI) and has received a total amount of 4.01 M€ for the period 2007–2014. European (FEDER), National (through Ministry of research, Inria, Inserm and ANR) and local councils (Brittany Region, Ille et Vilaine, and Rennes Metropole) have joined their effort to support this operation for a total amount of 4 010 k€ (600 k€ for the infrastructures, 2 850 k€ for the equipments and 560 k€ for the functioning). This application was set up through the Regional PIMATGI initiative coordinated by INSERM in Brittany (C. Roux). The overall PIMATGI initiative served for the financing of three distinct, but complementary, platforms: Neurinfo, TheraFONC as a technical platform dedicated to therapy guided by functional imaging especially in the oncology domain (Inserm U650 - LaTIM, Dir. Ch. Roux, Brest), and TherA-Image as a platform dedicated to image guided mini-invasive surgery and therapy especially in the domain of cardio-vascular diseases (U642 -LTSI, Dir. L. Senhadji, Rennes).

Concerning the Neurinfo Platform, the activity domain is a continuum between methodological and technological research built around specific clinical research projects. The ambition is to do innovation in science, technology and medical technology transfer for the implementation on the clinical field. On the medical field, the translational research domain mainly concerns medical imaging and more specifically the clinical neurosciences. Among them are multiple sclerosis, epilepsy, neurodegenerative, neurodevelopmental and psychiatric diseases, surgical procedures of brain lesions, neuro-oncology and radiotherapy planning. Beyond these CNS applications, the platform is also open to alternative applications. Neurinfo ambitions to support the emergence of research projects based on their level of innovation, their pluri-disciplinarity and their ability to foster collaborations between different actors (public and private research entities, different medical specialties, different scientific profiles).

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

In this context, a new research 3T MRI system (Siemens Verio system) was acquired in summer 2009 in order to develop the clinical research in the domain of morphological, functional, structural and cellular in-vivo imaging. In 2014 a new equipment for simultaneous recording of EEG and MRI images has been acquired from Brain Product. In 2015, a mock scanner for experimental set-up has been acquired as well as a new High Performance Computing environment made of one large computing cluster and a data center that is shared and operated by the Inria center at IRISA (UMR CNRS 6074). The computation cluster (240 cores) and the data center (up to 50 TB) are dedicated to host and process imaging data produced by the Neurinfo platform, but also by other research partners that share their protocols on the Neurinfo neuroinformatics system (currently more than 30 sites).

VisAGeS and its partners in the Neurinfo project are committed to use this new research platform for developing new regional, national and international collaborations around fundamental and applied clinical research projects dealing with in-vivo medical imaging.

In 2016, VisAGeS has been awarded by IBISA as a “Plateforme d’excellence”.

7. New Results

7.1. Image Computing: Detection, Segmentation, Registration and Analysis

7.1.1. Quantitative analysis of T2/T2 relaxation time alteration*

Participants: Benoit Combès, Anne Kerbrat, Olivier Commowick, Christian Barillot.

T2 and T2* relaxometric data [0] becomes a standard tool for the quantitative assessment of brain tissues and of their changes along time or after the infusion of a contrast agent. Being able to detect significant changes of T2/T2* relaxation time is an important issue. Generally, such a task is performed by comparing the variability level in the regions of interest to the variability in the normal appearance white matter. However, in the case of T2 and T2* relaxometry, this solution is highly problematic. Indeed the level of noise in the normal appearance white matter is significanlty smaller than the level of noise in more intense region (e.g. MS lesions). Our aim is to provide a Bayesian analysis of T2/T2* relaxometry estimation and alteration. More specifically, we build posterior distributions for the relaxation time and the relaxation offset by elucidating the dedicated Jeffreys priors. Then the resulting posterior distributions can be evaluated using a Monte Carlo Markov Chain algorithm. Such an analysis has three main advantages over the classical estimation procedure. First it allows in a simple way to compute many estimators of the posterior including the mode, the mean, the variance and confidence intervals. Then, it allows to include prior information. Finally, becauce one can extract confidence interval from the posterior, testing properly whether the true relaxometry time is included within a certain range of value given a confidence level is simple. This work was published as a conference paper in MICCAI 2016 [22].

7.1.2. Block-Matching Distortion Correction of Echo-Planar Images with Opposite Phase Encoding Directions

Participants: Renaud Hédouin, Olivier Commowick, Élise Bannier, Christian Barillot.

By shortening the acquisition time of MRI, Echo Planar Imaging (EPI) enables the acquisition of a large number of images in a short time, compatible with clinical constraints as required for diffusion or functional MRI. However such images are subject to large, local distortions disrupting their correspondence with the underlying anatomy. The correction of those distortions is an open problem, especially in regions where large deformations occur. We have proposed a new block-matching registration method to perform EPI distortion correction based on the acquisition of two EPI with opposite phase encoding directions (PED). It relies on new transformations between blocks adapted to the EPI distortion model, and on an adapted optimization scheme to ensure an opposite symmetric transformation. We have produced qualitative and quantitative results of the block-matching correction using different metrics on a phantom dataset and on in-vivo data. We have shown the ability of the block-matching to robustly correct EPI distortion even in strongly affected areas. This work has been accepted for publication in IEEE Transactions in Medical Imaging 2017.

[0] https://en.wikipedia.org/wiki/Relaxation_(NMR)

7.1.3. An a contrario approach for the detection of patient-specific brain perfusion abnormalities with arterial spin labelling

Participants: Pierre Maurel, Jean-Christophe Ferré, Christian Barillot.

In this work, we introduce a new locally multivariate procedure to quantitatively extract voxel-wise patterns of abnormal perfusion in individual patients. This a contrario approach uses a multivariate metric from the computer vision community that is suitable to detect abnormalities even in the presence of closeby hypo- and hyper-perfusions. This method takes into account local information without applying Gaussian smoothing to the data. Furthermore, to improve on the standard a contrario approach, which assumes white noise, we introduce an updated a contrario approach that takes into account the spatial coherency of the noise in the probability estimation. Validation is undertaken on a dataset of 25 patients diagnosed with brain tumors and 61 healthy volunteers. We show how the a contrario approach outperforms the massively univariate General Linear Model usually employed for this type of analysis. This work as been published in Neuroimage [14].

7.1.4. Dictionary Learning for Pattern Classification in Medical Imaging: Why Does Size Matter?

Participants: Hrishikesh Deshpande, Pierre Maurel, Christian Barillot.

Sparse representation based dictionary learning (DL) technique has proved to be an effective tool for image classification. While standard DL methods are effective in data representation, several discriminative DL methods have been proposed for learning dictionaries better suited for classification. Majority of these methods, in pattern recognition applications, learn the dictionaries for each class and com- pare the error terms of sparse reconstruction for each dictionary. However this raises a question that is still an open problem in the sparsity community: What role does the size of each dictionary play in the classification process? In this work, we prove that this parameter is pivotal, especially in cases where there are variability differences between classes. We illustrate our assertion on standard and discriminative DL techniques in two applications: Lips detection in face images and the classification of multiple sclerosis lesions in multi-channel brain MR images.

7.2. Image processing on Diffusion Weighted Magnetic Resonance Imaging

7.2.1. Maximum Likelihood Estimators of Brain White Matter Microstructure

Participant: Olivier Commowick.

Diffusion MRI is a key in-vivo non invasive imaging capability that can probe the microstructure of the brain. However, its limited resolution requires complex voxelwise generative models of the diffusion. Diffusion Compartment (DC) models divide the voxel into smaller compartments in which diffusion is homogeneous. We developed a comprehensive framework for maximum likelihood estimation (MLE) of such models that jointly features ML estimators of (i) the baseline MR signal, (ii) the noise variance, (iii) compartment proportions, and (iv) diffusion-related parameters. ML estimators are key to providing reliable mapping of brain microstructure as they are asymptotically unbiased and of minimal variance. We compare our algorithm (which efficiently exploits analytical properties of MLE) to alternative implementations and a state-of-the-art strategy. Simulation results show that our approach offers the best reduction in computational burden while guaranteeing convergence of numerical estimators to the MLE. In-vivo results also reveal remarkably reliable microstructure mapping in areas as complex as the centrum semiovale. Our ML framework accommodates any DC model and is available freely for multi-tensor models as part of the ANIMA software. This work was published as a conference paper in MICCAI 2016 [24].

7.3. EEG and MR Imaging

7.3.1. Multi-Modal EEG and fMRI Source Localization using Sparse Constraints

Participants: Saman Noorzadeh, Pierre Maurel, Christian Barillot.

In this work a multi-modal approach is introduced to estimate the brain neuronal sources based on EEG and fMRI. These two imaging techniques can provide complementary information about the neuronal activities of the brain. Each of these data modalities are first modeled linearly based on the sources. The sources are then estimated with a high spatio-temporal resolution based on a symmetrical integrated approach of these models. For a better estimation, a sparse constraint is also applied to the method based on the physiological knowledge that we have about the brain function. The results which are validated on the real data, shows the reconstruction of neuronal sources with the high spatio-temporal resolution. This is a joint work with Remi Gribonval.

7.3.2. Unimodal versus bimodal EEG-fMRI neurofeedback of a motor imagery task

Participants: Lorraine Perronnet, Marsel Mano, Élise Bannier, Christian Barillot.

In the context of the HEMISFER project, we proposed a simultaneous EEG-fMRI experimental protocol in which 10 healthy participants performed a motor-imagery task in unimodal and bimodal neurofeedback conditions. With this protocol we were able to compare for the first time the effects of unimodal EEG-neurofeedback and fMRI-neurofeedback versus bimodal EEG-fMRI-neurofeedback by looking both at EEG and fMRI activations. We also introduced a new feedback metaphor for bimodal EEG-fMRI-neurofeedback that integrates both EEG and fMRI signal in a single bi-dimensional feedback (a ball moving in 2D). Such a feedback is intended to relieve the cognitive load of the subject by presenting the bimodal neurofeedback task as a single regulation task instead of two. Additionally, this integrated feedback metaphor gives flexibility on defining a bimodal neurofeedback target. Participants were able to regulate activity in their motor regions in all neurofeedback conditions. Moreover, motor activations as revealed by offline fMRI analysis were stronger during EEG-fMRI-neurofeedback than during EEG-neurofeedback. This result suggests that EEG-fMRI-neurofeedback could be more specific or more engaging than EEG-neurofeedback. Our results also suggest that during EEG-fMRI-neurofeedback, participants tended to regulate more the modality that was harder to control. Taken together our results shed light on the specific mechanisms of bimodal EEG-fMRI-neurofeedback and on its added-value as compared to unimodal EEG-neurofeedback and fMRI-neurofeedback.

This work was done in collaboration with the Inria Hybrid and Athena teams. Experiments were conducted at the Neurinfo MRI research facility from University of Rennes 1. This was presented during the poster session of the 2016 Organization for Human Brain Mapping (OHBM) conference.

7.3.3. Brain training with Neurofeedback

Participants: Lorraine Perronnet, Christian Barillot.

We published a book chapter called Brain training with Neurofeedback in the book “Brain Computer Interfaces 1: Methods and Perspectives” (published in French and English) [26]. The first section of the chapter defines the concept of neurofeedback and gives an overall view of the current status in this domain. The second section describes the design of a NF training program and the typical course of a NF session, as well as the learning mechanisms underlying NF. The third section retraces the history of NF, explaining the origin of its questionable reputation and providing a foothold for understanding the diversity of existing approaches. The fourth section discusses how the fields of NF and BCIs might potentially overlap in future with the development of "restorative" BCIs. Finally, the fifth and last section presents a few applications of NF and summarizes the state of research of some of its major clinical applications.

7.3.4. Design of an Experimental Platform for Hybrid EEG-fMRI Neurofeedback Studies

Participants: Marsel Mano, Élise Bannier, Lorraine Perronnet, Christian Barillot.

During a neurofeedback (NF) experiment one or more brain activity measuring technologies are used to estimate the changes of the acquired neural signals that reflect the changes of the subject's brain activity in real-time. There exist a variety of NF research applications that use only one type of neural signals (i.e. uni-modal) like EEG or fMRI, but there are very few NF researches that use two or more neural signals (i.e. multi-modal). This is primarily because of the associated technical burdens.

We have developed, installed and successfully conducted used a hybrid EEG-fMRI platform for bi-modal NF experiments, as part of the project Hemisfer. Our system is based on the integration and the synchronization of an MR-compatible EEG and fMRI acquisition subsystems. The EEG signals are acquired with a 64 channel MR-compatible solution from Brain Products and the MR imaging is performed on a 3T Verio Siemens scanner (VB17) with a 12-ch head coil. We have developed two real-time pipelines for EEG and fMRI that handle all the necessary signal processing, the Joint NF module that calculates and fuses the NF and a visualize module that displays the NF to the subject. The control and the synchronization of both subsystems with each-other and with the experimental protocol is handled by the NF Control.

Our platform showed very good real-time performance with various pre-processing, filtering, and NF estimation and visualization methods. The entire fMRI process from acquisition to NF takes always less than 200ms, well below the TR of regular EPI sequences (2s). The same process for EEG, with NF update cycles varying 2-5Hz, is done in virtually real time (50Hz). Various NF tasks scenarios for regulating the measured brain activity were tested with subjects. In particular, the platform was used for a NF study on 10 subjects with over 50 sessions using three NF protocols based on motor imagery related brain activity: a) fMRI-NF, b) EEG-NF and c) EEG and fMRI-NF; and two online brain activity regulating protocols without NF. Our hybrid EEG-fMRI NF platform has been a very reliable environment for the NF experiments in our project. Its modular architecture is easily adaptable to different experimental environments, and offers high efficiency for optimal real-time NF applications.

7.4. Applications in Neuroradiology and Neurological Disorders

7.4.1. Imaging biomarkers in Multiple Sclerosis: from image analysis to population imaging

Participants: Christian Barillot, Gilles Edan, Olivier Commowick.

The production of imaging data in medicine increases more rapidly than the capacity of computing models to extract information from it. The grand challenges of better understanding the brain, offering better care for neurological disorders, and stimulating new drug design will not be achieved without significant advances in computational neuroscience. The road to success is to develop a new, generic, computational methodology and to confront and validate this methodology on relevant diseases with adapted computational infrastructures. This new concept sustains the need to build new research paradigms to better understand the natural history of the pathology at the early phase; to better aggregate data that will provide the most complete representation of the pathology in order to better correlate imaging with other relevant features such as clinical, biological or genetic data. In this context, one of the major challenges of neuroimaging in clinical neurosciences is to detect quantitative signs of pathological evolution as early as possible to prevent disease progression, evaluate therapeutic protocols or even better understand and model the natural history of a given neurological pathology. Many diseases encompass brain alterations often not visible on conventional MRI sequences, especially in normal appearing brain tissues (NABT). MRI has often a low specificity for differentiating between possible pathological changes which could help in discriminating between the different pathological stages or grades. The objective of medical image analysis procedures is to define new quantitative neuroimaging biomarkers to track the evolution of the pathology at different levels. We have published a position paper in Medical Image Analysis [2] that illustrates this issue in one acute neuro-inflammatory pathology: Multiple Sclerosis (MS). It exhibits the current medical image analysis approaches and explains how this field of research will evolve in the next decade to integrate larger scale of information at the temporal, cellular, structural and morphological levels.

7.4.2. Multiple Sclerosis lesion segmentation using an automated multimodal Graph Cut

Participants: Jérémy Beaumont, Olivier Commowick, Christian Barillot.

In this work, we present an algorithm for Multiple Sclerosis (MS) lesion segmentation. Our method is fully automated and includes three main steps: 1. the computation of a rough total lesion load in order to optimize the parameter set of the following step; 2. the detection of lesions by graph cut initialized with a robust Expectation-Maximization (EM) algorithm; 3. the application of rules to remove false positives and to adjust the contour of the detected lesions. This work was part of the FLI 2016 MSSEG challenge data organized at MICCAI 2016 [25].

7.4.3. Automatic Multiple Sclerosis lesion segmentation from Intensity-Normalized multi-channel MRI

Participants: Jérémy Beaumont, Olivier Commowick, Christian Barillot.

In the context of the FLI MICCAI 2016 MSSEG challenge for lesion segmentation, we present a fully automated algorithm for Multiple Sclerosis (MS) lesion segmentation. Our method is composed of three main steps. First, the MS patient images are registered and intensity normalized. Then, the lesion segmentation is done using a voxel-wise comparison of multi-channel Magnetic Resonance Images (MRI) against a set of controls. Finally, the segmentation is refined by applying several lesion appearance rules. This work was part of the FLI 2016 MSSEG challenge data organized at MICCAI 2016 [21].

7.5. Management of Information in Neuroimaging

Participants: Michael Kain, Olivier Commowick, Élise Bannier, Inès Fakhfakh, Justine Guillaumont, Florent Leray, Yao Chi, Christian Barillot.

The major topic that is addressed in this period concern the sharing of data and processing tools in neuroimaging (through the "Programme d'Investissement d'Avenir" project such as OFSEP and FLI-IAM) which led to build a suitable architecture to share images and processing tools, started from the NeuroBase project (supported by the French Ministry of Research). Our overall goal within these projects is to set up a computer infrastructure to facilitate the sharing of neuroimaging data, as well as image processing tools, in a distributed and heterogeneous environment. These consortium gathered expertise coming from several complementary domains of expertise: image processing in neuroimaging, workflows and GRID computing, ontology development and ontology-based mediation. This enables a large variety of users to diffuse, exchange or reach neuroimaging information with appropriate access means, in order to be able to retrieve information almost as easily as if the data were stored locally by means of the "cloud computing" Storage as a Service (SaaS) concept As an example, the Shanoir environment has been succesfully deployed to the Neurinfo platform were it is routinely used to manage images of the research studies. It is also currently being deployed for two large projets: OFSEP ("Observatoire Français de la Sclérose en Plaques") where up to 30000 patients will be acquired on a ten years frame, and the Image Analysis and Mangment (IAM) node of the France Life Imaging national infrastructure (FLI-IAM). Our team fulfills multiple roles in this nation-wide FLI project. Christian Barillot is the chair of the IAM node, Olivier Commowick is participating in the working group workflow and image processing and Michael Kain is the technical manager of the node. Apart from the team members, software solutions like medInria and Shanoir are part of the final software platform.

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Contracts with Industry

8.1.1. Siemens

duration: 5 years from 2011/10/26, extended until end of 2017.

In the context of the Neurinfo imaging platform, a partnership between Siemens SAS - Healthcare and University of Rennes 1 was signed in October 2011 for 5 years. This contract defines the terms of the collaboration between Siemens and the Neurinfo platform. The Neurinfo platform has received work in progress (WIP) sequences from Siemens in the form of object code for evaluation in the context of clinical research. The Neurinfo platform has also received source code of selected MRI sequences. This is a major advance in the collaboration since it will enable the development of MRI sequences on site.

8.2. Bilateral Grants with Industry

8.2.1. MEDday

As part of its activities, MEDday led the final testing phase on patients diagnosed from Multiple Sclerosis in order to find treatment of progressive multiple sclerosis. This is done in partnership with several hospitals in France. The goal is to achieve an effective treatment for this disease. The role of the team in this industrial grant is to develop new algorithms to perform the processing and the analysis of the images from this study.

9. Partnerships and Cooperations

9.1. Regional Initiatives

9.1.1. Biogenouest

The VisAGeS team and the Neurinfo platform integrated the Biogenouest "Groupement d'Intérêt Scientifique (GIS)" in 2012. Biogenouest is a Western France life science and environment core facility network. Research programmes are undertaken in the fields of Marine biology, Agriculture/Food-processing, Human health, and Bioinformatics. Set up in keeping with the inter-regional principle of complementarity, Biogenouest coordinates over twenty technological core facilities in both the Brittany and Pays de la Loire regions.

9.1.2. Projet Fondation de France: PERINE

Participants: Élise Bannier, Isabelle Corouge, Olivier Commowick, Jean-Christophe Ferré, Christian Barillot.

This study evaluates the effect of prenatal exposure to neurotoxicants on the developing brain. Following previous studies in the PELAGIE cohort this MRI study involves ASL, Diffusion and working memory as well as motor inhibition BOLD fMRI together with neuropsychological tests in children. Inclusions have started in November 2014 and lasted for 2 years. The MRI acquisitions of the PERINE projects have all been performed and 101 children included. A post-doc will start in April 2017 to process the diffusion MRI and ASL data of this project.

9.1.3. Fondation de l'Avenir - Stroke, rehabilitation and fMRI

Participants: Élise Bannier, Isabelle Bonan, Isabelle Corouge, Jean-Christophe Ferré, Christian Barillot, Jean-Yves Gauvrit.

duration: 12 months from November 2012. Project extended in 2015.
A complementary funding (20 000€) was obtained to support a new research project on rehabilitation of stroke patients. The fMRI tasks were setup and validated on healthy controls (paper ready for submission). The project was extended in 2014 to recruit more patients.

9.1.4. Projet Fondation de France: EPMR-MA

Participants: Pierre-Yves Jonin, Élise Bannier, Christian Barillot, Isabelle Corouge, Quentin Duché, Jean-Christophe Ferré.

This project evaluates memory effects in healthy adults and in patients presenting cognitive impairments using BOLD fMRI and diffusion MRI. A pilot study has been completed in 2015 in order to optimize the experimental design. The inclusions of patients started in 2016 and are ongoing. A Post Doc was recruited to work on fMRI and DTI processing.

9.1.5. Allocation d'Installation Scientifique – Rennes Métropole

Participant: Emmanuel Caruyer.

Diffusion MRI has been a tremendous tool for the diagnosis of a number of brain pathologies such as abnormal development, neuro-degenerative or inflammatory disorders or brain tumors. Typical resolution in diffusion MRI is about 2mm – this suggests that in white matter, any volume element may contain millions of axons. Although currently we can characterize molecular diffusion, recent developments in diffusion MRI have shown the possibility to quantify more specifically some physical tissue parameters in white matter, such as axonal density and diameter: this means that we can retrieve information from a much smaller scale than the typical imaging resolution.

Acquisition time for this kind of measurements remains long and largely incompatible with in vivo application in humans. This projects aims at developing novel signal processing and acquisition methods for the reconstruction of microstructural informations in a reasonable acquisition time. We will study how sparse representations can be applied to the diffusion signal, in order to enable microstructure information reconstruction. In conjunction with this, we will develop acquisition sequences adapted to these sparse represnetations, in order to reconstruct the diffusion signal from fewer measurements, using results from the compressive sensing theory.

9.2. National Initiatives

9.2.1. ANR

9.2.1.1. ANR "MAIA", 2015 generic projects program

Participants: Maia Proisy, Pierre Maurel, Antoine Legouhy, Olivier Commowick, Isabelle Corouge, Jean-Christophe Ferré, Christian Barillot.

Each year in France, 55 000 children are born prematurely, i.e., before the 37th week of gestation. Long-term studies of the outcome of prematurely born infants have clearly documented that the majority of such infants may have significant motor, cognitive, and behavioral deficits.

However, there is a limited understanding of the nature of the cerebral abnormality underlying these adverse neurologic outcomes. In this context, the emergence of new modalities of 3D functional MRI, e.g., Arterial Spin Labeling (ASL), or optical imaging technologies, e.g., Near InfraRed Spectroscopy (NIRS), brings new perspectives for extracting cognitive information, via metabolic activity measures. Other classical technics devoted to cerebral signal measurement, such as ElectroEncephaloGraphy (EEG), provide cognitive information at the cortical level. Each of these various non-invasive imaging technologies brings substantial and specific information for the understanding of newborn brain development.

This project aims at developing innovative approaches for multi-image / multi-signal analysis, in order to improve neurodevelopment understanding methods. From a fundamental point of view, mathematics and computer science have to be considered in association with imaging physics and medicine, to deal with open issues of signal and image analysis from heterogeneous data (image, signal), considered in the multiphysics contexts related to data acquisition (magnetic, optic, electric signals) and biophysics modeling of the newborn brain. A sustained synergy between all these scientific domains is then necessary.

Finally, the sine qua non condition to reach a better understanding of the coupled morphological- cognitive development of premature newborns, is the development of effective software tools, and their distribution to the whole medical community. The very target of this project will be the design of such software tools for medical image / signal analysis, actually operational in clinical routine, and freely available. Academic researchers and industrial partners will work in close collaboration to reach that ambitious goal.

9.2.2. Competitivity Clusters

9.2.2.1. The HEMISFER Project

Participants: Élise Bannier, Jean-Marie Batail, Isabelle Bonan, Isabelle Corouge, Jean-Christophe Ferré, Jean-Yves Gauvrit, Pierre Maurel, Lorraine Perronnet, Christian Barillot.

The HEMISFER project ("Hybrid Eeg-MrI and Simultaneous neuro-FEedback for brain Rehabilitation") will be conducted at Inria Rennes with the support of the Cluster of Excellence "CominLabs" [0]. The goal of HEMISFER is to make full use of the neurofeedback paradigm in the context of rehabilitation and psychiatric disorders. The major breakthrough will come from the use of a coupling model associating functional and metabolic information from Magnetic Resonance Imaging (fMRI) to Electro-encephalography (EEG) to "enhance" the neurofeedback protocol. We propose to combine advanced instrumental devices (Hybrid EEG and MRI platforms), with new man-machine interface paradigms (Brain computer interface and serious gaming) and new computational models (source separation, sparse representations and machine learning) to provide novel therapeutic and neuro-rehabilitation paradigms in some of the major neurological and psychiatric disorders of the developmental and the aging brain (stroke, attention-deficit disorder, language disorders, treatment-resistant mood disorders, ...). This project will be conducted with the HYBRID and PANAMA Teams from Inria Rennes, the EA 4712 team from University of Rennes I and the ATHENA team from Inria Sophia-Antipolis. This work will benefit from the research 3T MRI and MRI-compatible EEG systems provided by the NeurInfo in-vivo neuroimaging platform on which these new research protocols will be set up. A budget of 500keuros will be provided by the CominLabs cluster in the next 3 years to support this project (through experimental designs, PhDs, Post-docs and Expert Engineers).

9.2.2.2. France Life Imaging (FLI)

Participants: Christian Barillot, Olivier Commowick, Michael Kain, Florent Leray, Julien Louis, Aneta Morawin, Mathieu Simon, Yao Chi.

France Life Imaging (FLI) is a proposed large-scale research infrastructure project aimed at establishing a coordinated and harmonized network of biomedical imaging in France. This project was recently selected by the call “Investissements d’Avenir - Infrastructure en Biologie et Santé”. One node of this project is the node Information Analysis and Management (IAM), a transversal node build by a consortium of teams that will contribute to the construction of a network for data storage and information processing. Instead of building yet other dedicated facilities, the IAM node will use already existing data storage and information processing facilities (LaTIM Brest; CREATIS Lyon; CIC-IT Nancy; VisAGeS U746 Inria Rennes; CATI CEA Saclay; LSIIT/ICube Strasbourg) that will increase their capacities for the FLI infrastructure. Inter-connections and access to services will be achieved through a dedicated software platform that will be developed based on the expertise gained through successful existing developments. The IAM node has several goals. It aims first at building a versatile facility for data management that will inter-connect the data production sites and data processing for which state-of-the-art solutions, hardware and software, will be available to infrastructure users. Modular solutions are preferred to accommodate the large variety of modalities acquisitions, scientific problems, data size, and adapted for future challenges. Second, it aims at offering the latest development that will be made available to image processing research teams. The team VisAGeS fulfills multiple roles in this nation-wide project. Christian Barillot is the chair of the node IAM, Olivier Commowick is participating in the working group workflow and image processing and Michael Kain the technical manager. Apart from the team members, software solutions like medInria and Shanoir will be part of the final software platform.

9.2.2.3. OFSEP

Participants: Justine Guillaumont, Élise Bannier, Christian Barillot, Olivier Commowick, Gilles Edan, Jean-Christophe Ferré, Michael Kain, Inès Fakhfakh.

The French Observatory of Multiple Sclerosis (OFSEP) is one of 10 projects selected in January 2011 in response to the call for proposal in the “Investissements d’Avenir - Cohorts 2010” program launched by the French Government. It allows support from the National Agency for Research (ANR) of approximately € 10

[0] https://iww.inria.fr/cominlabs-newsletter/april-2013-four-projects-selected/#hemisfer

million for 10 years. It is coordinated by the Department of Neurology at the Neurological Hospital Pierre Wertheimer in Lyon (Professor Christian Confavreux), and it is supported by the EDMUS Foundation against multiple sclerosis, the University Claude Bernard Lyon 1 and the Hospices Civils de Lyon. OFSEP is based on a network of neurologists and radiologists distributed throughout the French territory and linked to 61 centers. OFSEP national cohort includes more than 50,000 people with Multiple Sclerosis, approximately half of the patients residing in France. The generalization of longitudinal monitoring and systematic association of clinical data and neuroimaging data is one of the objectives of OFSEP in order to improve the quality, efficiency and safety of care and promote clinical, basic and translational research in MS. For the concern of data management, the Shanoir platform of Inria has been retained to manage the imaging data of the National OFSEP cohort in multiple sclerosis.

9.2.2.4. PHRC EMISEP: Evaluation of early spinal cord injury and late physical disability in Relapsing Remitting Multiple Sclerosis

Participants: Élise Bannier, Christian Barillot, Emmanuel Caruyer, Benoit Combès, Olivier Commowick, Gilles Edan, Jean-Christophe Ferré, Anne Kerbrat, Haykel Snoussi.

Multiple Sclerosis (MS) is the most frequent acquired neurological disease affecting young adults (1/1000 inhabitants in France) and leading to impairment. Early and well adapted treatment is essential in patients presenting aggressive forms of MS. This PHRC project focusses on physical impairment and especially on the ability to walk. Several studies, whether epidemiolgic or based on brain MRI, have shown that several factors were likely to announce aggressive development of the disease, such as age, number of focal lesions on baseline MRI, clinical activity. However, these factors only partially explain physical impairment progression, preventing their use at the individual level. Spinal cord is often affected in MS, as demonstrated in postmortem or imaging studies. Yet, early radiological depiction of spinal cord lesions is not always correlated with clinical symptoms. Preliminary data, on reduced number of patients, and only investigating the cervical spinal cord have shown that diffuse spinal cord injury, observed via diffusion or magnetisation transfer imaging, would be correlated with physical impairment as evaluated by the EDSS score. Besides, the role of early spinal cord affection (first two years) in the evolution of physical impairment remains unknown.

In this project, we propose to address these different issues and perform a longitudinal study on Relapsing Remitting Multiple Sclerosis (RRMS) patients, recruited in the first year of the disease. Our goal is to show that diffuse and focal lesions detected spinal cord MRI in the first 2 years can be used to predict disease evolution and physical impairment at 5 years. Twelve centers are involved in the study to include 80 patients.

To date, 65 of the 80 subjects have been included. H. Snoussi is working in the scope of his PhD thesis on diffusion imaging in the spinal cord and has dedicated his first year to literature review and definition of methodological aspects to tackle starting with distortion correction. B. Combès started as a post doc in November 2016 to process the EMISEP imaging data, starting with morphological data processing (registration, segmentation) and magnetization transfer data processing.

9.3. European Initiatives

9.3.1. FP7 & H2020 Projects

9.3.1.1. EuroBioimaging

Type: CAPACITIES

Challenge: Provide access and training in imaging technologies, and share the best practice and image data in order to make Euro-BioImaging an engine that will drive European innovation in imaging research and technologies

Instrument: Combination of COLLABORATIVE PROJECTS and COORDINATION and SUPPORT ACTIONS

Objective: Euro-BioImaging is a large-scale pan-European research infrastructure project on the

European Strategy Forum on Research Infrastructures (ESFRI) Roadmap.

Duration: December 2010 - 2016

Coordinators: Jan Ellenberg (EMBL) and Oliver Speck (University of Magdeburg)

Partner: EMBL (Germany); Erasmus Medical Center (Netherlands) for WG11

Inria contact: C. Kervrann, Christian Barillot

Abstract: Euro-BioImaging is a pan-European infrastructure project whose mission is to build a distributed imaging infrastructure across Europe that will provide open access to innovative biological and medical imaging technologies for European researchers. The project is funded by the EU and currently the consortium is finalizing the basic principles for the operation of future Euro-BioImaging organisation.

Euro-BioImaging will be governed by representatives of the European countries that will join Euro-BioImaging (Euro-BioImaging member states).

The infrastructure established by Euro-BioImaging will consist of a set of geographically distributed but strongly interlinked imaging facilities (Euro-BioImaging Nodes), which will be selected among the leading European imaging facilities based on an independent evaluation process.

Inria and the VisAGeS team is involved through the FLI national infrastructure and contributes to the WG11 Working Group on Data Storage and Analysis. This WG performs a series of tasks to define a European Biomedical Imaging Data Storage and Analysis infrastructure plan for the construction phase.

9.3.1.2. H2020 OpenAire-Connect

Program: E-INFRA

Topic: EINFRA-22-2016

Type of Action: RIA

Project acronym: OpenAIRE-Connect

Project title: OpenAIRE - CONNECTing scientific results in support of Open Science

Acceptation date: 01/09/2016

Open Science is around the corner. Scientists and organizations see it as a way to speed up, improve quality and reward, while policy makers see it as a means to optimize cost of science and leverage innovation. Open Science is an emerging vision, a way of thinking, whose challenges always gaze beyond its actual achievements. De facto, today's scientific communication ecosystem lacks tools and practices to allow researchers to fully embrace Open Science. OpenAIREConnect aims to provide technological and social bridges, and deliver services enabling uniform exchange of research artefacts (literature, data, and methods), with semantic links between them, across research communities and content providers in scientific communication. It will introduce and implement the concept of Open Science as a Service (OSaaS) on top of the existing OpenAIRE infrastructure, delivering out-of-the-box, on-demand deployable tools. OpenAIRE-Connect will adopt an end-user driven approach (via the involvement of 5 prominent research communities), and enrich the portfolio of OpenAIRE infrastructure production services with a Research Community Dashboard Service and a Catch-All Notification Broker Service. The first will offer publishing, interlinking, packaging functionalities to enable them to share and re-use their research artifacts (introducing methods, e.g. data,software, protocols). This effort, supported by the harvesting and mining "intelligence" of the OpenAIRE infrastructure, will provide communities with the content and tools they need to effectively evaluate and reproduce science. OpenAIRE-Connect will combine dissemination and training with OpenAIRE's powerful NOAD network engaging research communities and content providers in adopting such services. These combined actions will bring immediate and long-term benefits to scholarly communication stakeholders by affecting the way research results are disseminated, exchanged, evaluated, and re-used.

In this project VisAGeS is acting, through CNRS, as the French coordinator to develop the link with the Neuroimaging research community. This will be performed in the context of the FLI-IAM national infrastructure

9.3.2. Collaborations in European Programs, Except FP7 & H2020

9.3.2.1. Kic-EIT-eHealth

Program: KIC-EIT: European Institute of Innovation and Technology

Project acronym: e-Health

Project title: Innovation for healthy living and active ageing

Acceptation date: 01/12/2014

website: http://eithealth.eu/about-us/

EIT Health aims to promote entrepreneurship and develop innovations in healthy living and active ageing, providing Europe with new opportunities and resources. EIT Health will enable citizens to lead healthier and more productive lives by delivering products, services and concepts that will improve quality of life and contribute to the sustainability of healthcare across Europe. EIT Health is a strong, diverse and balanced partnership of best-in-class organisations in education, research, technology, business creation and corporate and social innovation. EIT Health intends to foster cooperation and unlock Europe's innovation and growth potential – developing and retaining the best talents, creating high-quality jobs and boosting the global competitiveness of European industry. VisAGeS is involved in this project through the Inserm and Inria institutions. Christian Barillot is representing Inria as one expert in the dedicated WG "Healthy Brain". VisAGeS is also concerned by the WG "big data".

9.4. International Initiatives

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

9.4.1.1. BARBANT

Title: Boston and Rennes, a Brain image Analysis Team

International Partner (Institution - Laboratory - Researcher):

Harvard University (United States) - Mathematics Department - Simon K. Warfield

Start year: 2015

See also: https://team.inria.fr/barbant/

BARBANT is an Inria associate team shared between Inria VisAGeS research team and the Computational Radiology Laboratory at the Boston Children's hospital (Harvard Medical School). This associate team aims at better understanding the behavior of normal and pathological Central Nervous System (CNS) organs and systems. Pathologies of particular interest to us are multiple sclerosis, psychiatric, and pediatric diseases such as pediatric multiple sclerosis or tuberous sclerosis. A major challenge is to characterize the future course of the pathological processes in each patient as early as possible in order to predict the progression of the disease and/or adverse neurological outcomes, and to develop better techniques for both monitoring response to therapy and for altering therapy (duration, dose and nature) in response to patient-specific changes in imaging characteristics. At term, this project will allow to introduce objective figures to correlate qualitative and quantitative phenotypic markers coming from the clinic and image analysis, mostly at the early stage of the pathologies. This will allow for the selection or adaptation of the treatment for patients at an early stage of the disease.

9.4.1.2. Informal International Partners

- Collaboration with Sherbrooke University (Sherbrooke, Canada): From Jun to Aug 2016, Michael Paquette, PhD student from Sherbrooke supervised by Maxime Descoteaux, visited the VisAGeS team to collaborate with Emmanuel Caruyer on the development on new analysis techniques for the structural brain connectome. This visit was funded by a MITACS/Inria scholarship.

- Collaboration with LTS5, EPFL (Lausanne, Switzerland) and Computer Science department, University of Verona (Verona, Italy): Alessandro Daducci, Gabriel Girard and Jean-Philippe Thiran visited the VisAGeS team for a 2 days workshop on the development of novel validation methods for the human brain connectome using software generated phantoms.
- Collaboration with the Mathematics department, Politecnico di Milano (Italy): Olivier Commowick and Christian Barillot visited the department for the annual meeting of the Italian statistical society and collaborated with Aymeric Stamm and Simone Vantini.
- Collaboration with the Microstructure Imaging Group, UCL (London, UK): Christian Barillot, Emmanuel Caruyer, Olivier Commowick and Sudhanya Chatterjee visited the group of Daniel Alexander for a workshop on “MRI based Virtual Histology: Meeting Tomorrow’s Healthcare Challenges Today”
- visit of Tobias Kober and Bénédicte Maréchal from the ACIT Siemens research group in Lausanne [0] to discuss potential collaborations on the MP2Rage sequence and other brain MR imaging topics

9.5. International Research Visitors

9.5.1. Visits of International Scientists

- Simon Warfield and Benoit Scherrer, Harvard University, visited the VisAGeS team for the annual seminar on Jun 9-10 2016.
- From Jun to Aug 2016, Michael Paquette, PhD student from Sherbrooke supervised by Maxime Descoteaux, visited the VisAGeS team to collaborate with Emmanuel Caruyer on the development on new analysis techniques for the structural brain connectome. This visit was funded by a MITACS/Inria scholarship.
- Alessandro Daducci, Gabriel Girard and Jean-Philippe Thiran visited the VisAGeS team for a 2 days workshop on the development of novel validation methods for the human brain connectome using software generated phantoms.

9.5.2. Visits to International Teams

- Sudhanya Chatterjee visited the Computational Radiology Lab, Boston Children’s Hospital, Harvard University for 3 weeks in Oct-Nov 2016. This stay was funded by the international program of University of Rennes 1. Christian Barillot, Emmanuel Caruyer and Olivier Commowick visited the same lab for a 3 days workshop in the context of the Associate Team.
- Christian Barillot, Emmanuel Caruyer, Olivier Commowick and Sudhanya Chatterjee visited the Microstructure Imaging Group, UCL (London, UK) of Daniel Alexander for a workshop on “MRI based Virtual Histology: Meeting Tomorrow’s Healthcare Challenges Today”
- Olivier Commowick and Christian Barillot visited the Mathematics department, Politecnico di Milano (Italy) for the annual meeting of the Italian statistical society and collaborated with Aymeric Stamm and Simone Vantini.

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events Organisation

10.1.1.1. General Chair, Scientific Chair

- Olivier Commowick was General Chair of the “Multiple Sclerosis Lesions Segmentation Challenge”, MICCAI 2016.

[0] http://w1.siemens.ch/home/ch/de/healthcare/produkte/ACIT/Pages/ACIT.aspx

10.1.1.2. Member of the Organizing Committees

- Olivier Commowick, Michael Kain, Florent Leray, Jean-Christophe Ferré, Anne Kerbrat, Mathieu Simon and Christian Barillot organized the “Multiple Sclerosis Lesions Segmentation Challenge”, MICCAI 2016.
- Christian Barillot is member of the Board of Directors of IPMI conference series (Information Processing in Medical Imaging)

10.1.2. Scientific Events Selection

10.1.2.1. Chair of Conference Program Committees

- Christian Barillot was area chair of SPIE Medical Imaging 2016
- Christain Barillot was area chair of IEEE ISBI 2016

10.1.2.2. Member of the Conference Program Committees

- Christian Barillot was TPC Member of PatchMI-2016, MICCAI-MCV2016
- Emmanuel Caruyer was Program Committee member of the CDMRI MICCAI workshop.

10.1.2.3. Reviewer

- ISBI (Emmanuel Caruyer, Olivier Commowick), ISMRM (Élise Bannier), MICCAI (Emmanuel Caruyer, Olivier Commowick).

10.1.3. Journal

10.1.3.1. Member of the Editorial Boards

- Christian Barillot is member of Editorial Boards of Medical Image Analysis, Current Medical Imaging Reviews, ISRN Signal Processing
- Christian Barillot is Editor-in-Chief of Frontiers in ICT: Computer Image Analysis.

10.1.3.2. Reviewer - Reviewing Activities

- Am J Neuroradiol (Élise Bannier), Comput Biol Med (Christian Barillot), Comput Meth Prog Bio (Christian Barillot), Front Neurosc (Pierre Maurel), Hum Brain Mapp (Emmanuel Caruyer), IEEE TMI (Pierre Maurel, Olivier Commowick), Med Image Anal (Olivier Commowick), Med Phys (Christian Barillot) Neuroimage (Christian Barillot, Isabelle Corouge, Emmanuel Caruyer, Olivier Commowick), Pattern Recog Lett (Christian Barillot).

10.1.4. Invited Talks

- Gilles Edan gave an invited keynote at the world 2016 ECTRIMS conference (London, UK)
- Christian Barillot published an invited position paper for the 20th anniversary of Medical Image Analysis
- Christian Barillot gave an invited lecture at the 25ème COLLOQUE DE LA CONFÉRENCE NATIONALE DES COMITÉS DE PROTECTION DES PERSONNES (CNCP)
- Christian Barillot gave an invited lecture at the Global Bioimaging Training Program, Eurobioimaging ESFRI program, EMBL, Germany
- Christian Barillot gave an invited lecture at the Miccai-BrainLes Workshop 2016, Athenes, GR
- Christian Barillot gave an invited lecture at the Maria de Maeztu Strategic Research Program; Department of Information and Communication Technologies, UPF, Barcelona, Spain
- Christian Barillot gave an invited lecture at the RIR 2016 - Emerging challenges in neuroscience, neurology & psychiatry, Paris, France
- Christian Barillot gave an invited lecture at the Biomedical Imaging Seminar, Erasmus MC, Rotterdam
- Christian Barillot gave an invited lecture at the FCRIN day on “Specificities of clinical research in imaging”

- Christian Barillot, Emmanuel Caruyer and Olivier Commowick gave an invited talk at the "MRI based Virtual Histology: Meeting Tomorrow's Healthcare Challenges Today" workshop, University College London, May 26-27th.

10.1.5. Leadership within the Scientific Community

- Gilles Edan was elected Fellow of the European Academy of Neurologie. Member of the EAN teaching committee in 2015
- Christian Barillot is member of the Scientific Council of the INS2I [0] Institute of CNRS since 2011 and is Chairman of the Board since 2015
- Christian Barillot is member of the C3N committee (CNRS)
- Christian Barillot is member of the scientific board of "GIS France Grilles"

10.1.6. Scientific Expertise

- Christian Barillot provided an expertise for the Royal Netherlands Academy of Arts and Sciences (KNAW)
- Christian Barillot provided expertise for the EPSRC, UK
- Christian Barillot provided an expertise for ANRT
- Christian Barillot provided expertise for the Assistant professor committee for the University of Paris Sud
- Emmanuel Caruyer provided expertise for the Inria Associate Team program.

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

Master SIBM, University of Angers-Brest-Rennes

Christian Barillot, Élise Bannier, Emmanuel Caruyer, Olivier Commowick, Isabelle Corouge, Jean-Yves Gauvrit, Sylvain Prima, 3D medical imaging – visualization, segmentation, fusion, management, normalization (26h)

Sylvain Prima, Master 1 SIBM, University of Rennes (5h)

Christian Barillot is responsible for one semester

Jean-Yves Gauvrit is the coordinator for the Master

École Supérieure d'Ingénieur de Rennes (ESIR): Pierre Maurel, General image processing (60h), Algorithmics and complexity (60h), Medical imaging (60h)

ENS Rennes: Pierre Maurel, Introduction to image processing (24h)

ISTIC – Université of Rennes 1: Emmanuel Caruyer, Software Engineering (12h)

10.2.2. Supervision

PhD: Hrishikesh Deshpande, "Dimensionality Reduction and Statistical Learning for Computational Modeling of Natural Evolution of Brain Pathologies", Inria, Jun 2012, Christian Barillot, Pierre Maurel.

PhD in progress: Sudhanya Chatterjee, "Image-based Tissue Compartment Characterization of Neural Circuits with in-vivo MRI", Inria, from Nov 2015, Christian Barillot, Olivier Commowick, Jean-Christophe Ferré, Simon Warfield.

PhD in progress: Renaud Hédouin, "Biomarker discovery in brain imaging by using diffusion MRI", Inria/Inserm, from November 2013, Christian Barillot, Olivier Commowick.

[0] http://csins2i.irisa.fr

PhD in progress: Cédric Meurée, “Quantitative Analysis Of Arterial Spin Labeling MRI For Robust Parametric Information Of Perfusion Maps”, Inria / Siemens, from Mar 2014, Christian Barillot, Pierre Maurel.

PhD in progress: Corentin Vallée, “Joint estimation of neuronal activation, resting-state and basal metabolism from Arterial Spin Labeling”, Christian Barillot, Isabelle Corouge, Pierre Maurel.

PhD in progress: Antoine Legouhy, “Analyse IRM multimodale pour l’e´tude du de´veloppement ce´re´bral chez le pre´mature´”, from Nov 2016, Christian Barillot, Olivier Commowick, François Rousseau.

PhD in progress: Lorraine Perronnet, “Neurofeedback Using Virtual Reality And Combining Eeg-Mri For Brain Rehabilitation”, Inria/CominLabs Hemisfer project, from Dec 2013, Christian Barillot, Maureen Clerc (Inria Sophia-Antipolis), Anatole Lecuyer (HYBRID project), Fabien Lotte (Inria Bordeaux)

PhD in progress: Haykel Snoussi, “Diffusion MRI detection of early occurring spine lesions in relapsing-remitting multiple sclerosis with late physical impairment”, from Nov 2015, Christian Barillot, Gilles Edan, Emmanuel Caruyer

PhD in progress: Pierre-Yves Jonin, “Relationships between context-free and context-rich memory: cognitive and neural subtrates”, Inria/Inserm/CNRS from Oct 2014, Christian Barillot (co-supervisor)

PhD in progress: Maia Proisy, “Perfusion in neonates and in pediatric diseases”, Univ. Rennes 1/CHRU Rennes, from Oct 2014, Jean-Christophe Ferré (supervisor)

PhD in progress: Anne Kerbrat, “Quantitative MR imaging in MS for Brain and Spine”, Univ. Rennes 1/CHRU Rennes, from Oct 2014, Gilles Edan (supervisor)

10.2.3. Juries

- Pierre Maurel, PhD committee Hrishikesh Despande, Inria, Rennes July 2016.
- Christian Barillot, PhD committees: Hrishikesh Despande, Inria, Rennes July 2016.
- Christian Barillot, PhD reviewer: Mehdi Hadj-Hamou, Inria, Sophia; Dec 2016; Ester Bron, Erasmus MC, NL, March 2016

10.3. Popularization

- Inria demonstration stand, Journées Françaises de Radiologie.

11. Bibliography

Publications of the year

Articles in International Peer-Reviewed Journal

[1] C. BARILLOT, E. BANNIER, O. COMMOWICK, I. COROUGE, A. BAIRE, I. FACKFACK, J. GUILLAUMONT, Y. YAO, M. KAIN.*Shanoir: Applying the Software as a Service Distribution Model to Manage Brain Imaging Research Repositories*, in "Frontiers in information and communication technologies", October 2016 [*DOI :* 10.3389/FICT.2016.00025], http://www.hal.inserm.fr/inserm-01404864.

[2] C. BARILLOT, G. EDAN, O. COMMOWICK.*Imaging biomarkers in Multiple Sclerosis: from image analysis to population imaging*, in "Medical Image Analysis", 2016 [*DOI :* 10.1016/J.MEDIA.2016.06.017], https://hal.inria.fr/hal-01333583.

[3] L. BEUZIT, P.-A. ELIAT, V. BRUN, J.-C. FERRÉ, Y. GANDON, E. BANNIER, H. SAINT-JALMES.*Dynamic contrast-enhanced MRI: Study of inter-software accuracy and reproducibility using simulated and clinical data*, in "Journal of Magnetic Resonance Imaging", 2016, vol. 43, n^{o} 6, p. 1288-300 [*DOI :* 10.1002/JMRI.25101], https://hal-univ-rennes1.archives-ouvertes.fr/hal-01255866.

[4] A. COUM, F. NOURY, E. BANNIER, K. BEGRICHE, B. FROMENTY, Y. GANDON, H. SAINT-JALMES, G. GAMBAROTA.*The effect of water suppression on the hepatic lipid quantification, as assessed by the LCModel, in a preclinical and clinical scenario*, in "Magnetic Resonance Materials in Physics, Biology and Medicine", 2016, vol. 29, n^{o} 1, p. 29-37 [*DOI :* 10.1007/S10334-015-0508-1], https://hal-univ-rennes1.archives-ouvertes.fr/hal-01237087.

[5] P.-M. DAVID, N. ICARD, J.-Y. GAUVRIT, M. VÉRIN, F. LE JEUNE.*Interest of multi-modal imaging in bilateral lesions of basal ganglia: A case report of a post-anoxic dystonia*, in "European Journal of Nuclear Medicine and Molecular Imaging", October 2016, vol. 43, n^{o} 11, p. 2098–2099 [*DOI :* 10.1007/S00259-016-3413-6], https://hal-univ-rennes1.archives-ouvertes.fr/hal-01371659.

[6] R. H. FICK, D. WASSERMANN, E. CARUYER, R. DERICHE.*MAPL: Tissue Microstructure Estimation Using Laplacian-Regularized MAP-MRI and its Application to HCP Data*, in "NeuroImage", July 2016, vol. 134, p. 365–385 [*DOI :* 10.1016/J.NEUROIMAGE.2016.03.046], https://hal.inria.fr/hal-01291929.

[7] R. GONSETTE, M. DEBOUVERIE, C. SINDIC, J.-C. FERRÉ, G. EDAN.*Pixantrone: a B-cell-depleting immunosuppressant for multiple sclerosis patients with active disease*, in "Multiple Sclerosis Journal", May 2016, vol. 22, n^{o} 6, p. 817–821 [*DOI :* 10.1177/1352458515601902], https://hal-univ-rennes1.archives-ouvertes.fr/hal-01324700.

[8] R. HÉDOUIN, O. COMMOWICK, E. BANNIER, B. SCHERRER, M. TAQUET, S. K. WARFIELD, C. BARILLOT.*Block-Matching Distortion Correction of Echo-Planar Images With Opposite Phase Encoding Directions*, in "IEEE Transactions on Medical Imaging", 2017, 10, Accepted for publication, available on line [*DOI :* 10.1109/TMI.2016.2646920], https://hal.inria.fr/hal-01436561.

[9] S. JAMALI, R. FAHED, J. GENTRIC, L. LETOURNEAU-GUILLON, H. RAOULT, F. BING, L. ESTRADE, T. N. NGUYEN, E. TOLLARD, J.-C. FERRÉ, D. IANCU, O. NAGGARA, M. CHAGNON, A. WEILL, D. ROY, J. FOX, D. F. KALLMES, J. RAYMOND.*Inter- and Intrarater Agreement on the Outcome of Endovascular Treatment of Aneurysms Using MRA*, in "American Journal of Neuroradiology", 2016, vol. 37, n^{o} 5, p. 879 - 884 [*DOI :* 10.3174/AJNR.A4609], https://hal.inria.fr/hal-01377988.

[10] A. KERBRAT, J.-C. FERRÉ, P. FILLÂTRE, T. RONZIÈRE, S. VANNIER, B. CARSIN-NICOL, S. LAVOUÉ, M. VÉRIN, J.-Y. GAUVRIT, Y. LE TULZO, G. EDAN.*Acute Neurologic Disorder from an Inhibitor of Fatty Acid Amide Hydrolase*, in "New England Journal of Medicine", 2016, vol. 375, n^{o} 18, p. 1717–1725 [*DOI :* 10.1056/NEJMOA1604221], https://hal-univ-rennes1.archives-ouvertes.fr/hal-01412967.

[11] A. KIANI, A. ESQUEVIN, N. LEPAREUR, P. BOURGUET, F. LE JEUNE, J.-Y. GAUVRIT.*Main applications of hybrid PET-MRI contrast agents: a review*, in "Contrast Media and Molecular Imaging", 2016, vol. 11, n^{o} 2, p. 92–98 [*DOI :* 10.1002/CMMI.1674], https://hal-univ-rennes1.archives-ouvertes.fr/hal-01239820.

[12] J. KNAPP, B. COMBES, G. UMHANG, S. AKNOUCHE, L. MILLON.*Could the domestic cat play a significant role in the transmission of Echinococcus multilocularis ? A study based on qPCR analysis of cat feces in a rural area in France*, in "Parasite", October 2016, vol. 23, n^{o} 5-6, p. 589 - 599, https://hal.archives-ouvertes.fr/hal-01397027.

[13] A. J. MARCHAND, P. SEGUIN, Y. MALLEDANT, M. TALEB, H. RAOULT, J. Y. GAUVRIT.*Revised CT angiography venous score with consideration of infratentorial circulation value for diagnosing brain death*, in "Annals of Intensive Care", 2016, vol. 6, 88 [*DOI :* 10.1186/S13613-016-0188-7], https://hal-univ-rennes1.archives-ouvertes.fr/hal-01373307.

[14] C. MAUMET, P. MAUREL, J.-C. FERRÉ, C. BARILLOT.*An a contrario approach for the detection of patient-specific brain perfusion abnormalities with arterial spin labelling*, in "NeuroImage", July 2016, vol. 134 [*DOI :* 10.1016/J.NEUROIMAGE.2016.03.054], http://www.hal.inserm.fr/inserm-01291748.

[15] C. NADEAU, A. KRUPA, J. PETR, C. BARILLOT.*Moments-Based Ultrasound Visual Servoing: From Mono to Multi-plane Approach*, in "IEEE Transactions on Robotics", 2017 [*DOI :* 10.1109/TRO.2016.2604482], https://hal.inria.fr/hal-01385661.

[16] M. PROISY, B. BRUNEAU, C. ROZEL, C. TREGUIER, K. CHOUKLATI, L. RIFFAUD, P. DARNAULT, J.-C. FERRÉ.*Arterial spin labeling in clinical pediatric imaging*, in "Diagnostic and Interventional Imaging", 2016, vol. 97, n^{o} 2, p. 151–158 [*DOI :* 10.1016/J.DIII.2015.09.001], https://hal-univ-rennes1.archives-ouvertes.fr/hal-01361497.

[17] M. PROISY, S. MITRA, C. URIA-AVELLANA, M. SOKOLSKA, N. J. ROBERTSON, F. LE JEUNE, J.-C. FERRÉ.*Brain Perfusion Imaging in Neonates: An Overview*, in "American Journal of Neuroradiology", 2016, vol. 37, n^{o} 10, p. 1766 - 1773, Article en open access [*DOI :* 10.3174/AJNR.A4778], https://hal.inria.fr/hal-01377989.

[18] I. RACT, A. POUJADE, B. CARSIN-NICOL, F. MOURIAUX, J.-C. FERRÉ.*Spinal cord involvement in posterior reversible encephalopathy syndrome (PRES)*, in "Journal de Neuroradiologie / Journal of Neuroradiology", 2016, vol. 43, n^{o} 1, p. 56 - 58 [*DOI :* 10.1016/J.NEURAD.2015.09.004], https://hal.inria.fr/hal-01377978.

[19] A. RAILLON MOULINET, L. CHOCHINA, S. C. LEPLAIDEUR, E. BANNIER, J.-C. FERRÉ, I. BONAN.*Cortical activity and egocentric representation: An fMRI study*, in "Annals of Physical and Rehabilitation Medicine", 2016, vol. 59 [*DOI :* 10.1016/J.REHAB.2016.07.162], https://hal.inria.fr/hal-01378303.

[20] H. RAOULT, H. REDJEM, R. BOURCIER, A. GAULTIER-LINTIA, B. DAUMAS-DUPORT, J.-C. FERRÉ, F. EUGÈNE, R. FAHED, B. BARTOLINI, M. PIOTIN, H. DESAL, J.-Y. GAUVRIT, R. BLANC.*Mechanical thrombectomy with the ERIC retrieval device: initial experience*, in "Journal of neurointerventional surgery", 2016 [*DOI :* 10.1136/NEURINTSURG-2016-012379], https://hal.inria.fr/hal-01377990.

International Conferences with Proceedings

[21] J. BEAUMONT, O. COMMOWICK, C. BARILLOT.*Automatic Multiple Sclerosis lesion segmentation from Intensity-Normalized multi-channel MRI*, in "Proceedings of the 1st MICCAI Challenge on Multiple Sclerosis Lesions Segmentation Challenge Using a Data Management and Processing Infrastructure - MICCAI-MSSEG", Athens, Greece, October 2016, http://www.hal.inserm.fr/inserm-01424802.

[22] B. COMBÈS, A. KERBRAT, O. COMMOWICK, C. BARILLOT.*A Bayesian Model to Assess T2 Values and Their Changes Over Time in Quantitative MRI*, in "19th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)", Athens, Greece, October 2016, p. 570 - 578 [*DOI :* 10.1007/978-3-319-46726-9_66], http://www.hal.inserm.fr/inserm-01349557.

[23] O. COMMOWICK, A. STAMM, S. VANTINI, S. K. WARFIELD.*Maximum Likelihood Estimators of Brain White Matter Microstructure*, in "48th Scientific Meeting of the Italian Statistical Society", Salerno, Italy, June 2016, http://www.hal.inserm.fr/inserm-01347200.

[24] A. STAMM, O. COMMOWICK, S. K. WARFIELD, S. VANTINI.*Comprehensive Maximum Likelihood Estimation of Diffusion Compartment Models Towards Reliable Mapping of Brain Microstructure*, in "19th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)", Athens, Greece, October 2016, p. 622 - 630, To appear [*DOI :* 10.1007/978-3-319-46726-9_72], http://www.hal.inserm.fr/inserm-01349556.

Conferences without Proceedings

[25] J. BEAUMONT, O. COMMOWICK, C. BARILLOT.*Multiple Sclerosis lesion segmentation using an automated multimodal Graph Cut*, in "Proceedings of the 1st MICCAI Challenge on Multiple Sclerosis Lesions Segmentation Challenge Using a Data Management and Processing Infrastructure – MICCAI-MSSEG", Athens, Greece, O. COMMOWICK, F. CERVENANSKY, R. AMELI (editors), October 2016, p. 1-8, http://www.hal.inserm.fr/inserm-01417378.

Scientific Books (or Scientific Book chapters)

[26] L. PERRONNET, A. LÉCUYER, F. LOTTE, M. CLERC, C. BARILLOT.*Brain training with neurofeedback*, in "Brain-Computer Interfaces 1", Wiley-ISTE, July 2016, https://hal.inria.fr/hal-01413424.

[27] L. PERRONNET, A. LÉCUYER, F. LOTTE, M. CLERC, C. BARILLOT.*Entraîner son cerveau avec le neurofeedback*, in "Les interfaces cerveau-ordinateur 1", M. CLERC, L. BOUGRAIN, F. LOTTE (editors), ISTE editions, July 2016, p. 277-292, https://hal.inria.fr/hal-01413408.

Books or Proceedings Editing

[28] O. COMMOWICK, F. CERVENANSKY, R. AMELI (editors). *MSSEG Challenge Proceedings: Multiple Sclerosis Lesions Segmentation Challenge Using a Data Management and Processing Infrastructure*, 2016, http://www.hal.inserm.fr/inserm-01397806.

Other Publications

[29] M. MANO, E. BANNIER, L. PERRONNET, A. LÉCUYER, C. BARILLOT.*Design of an Experimental Platform for Hybrid EEG-fMRI Neurofeedback Studies*, June 2016, 22nd Annual Meeting of the Organization for Human Brain Mapping (OHBM 2016), Poster, https://hal.archives-ouvertes.fr/hal-01426072.

[30] C. MERCK, I. COROUGE, P.-Y. JONIN, B. DESGRANGES, J.-Y. GAUVRIT, S. BELLIARD.*Does the left posterior fusiform gyrus play a critical role in fruit and vegetables categorization? Evidence from 19 semantic dementia patients*, July 2016, International Neuropsycological Society Mid-Year Meeting, Poster [*DOI :* 10.13140/RG.2.2.24681.70242], http://www.hal.inserm.fr/inserm-01417028.

[31] L. PERRONNET, A. LÉCUYER, M. MANO, E. BANNIER, F. LOTTE, M. CLERC, C. BARILLOT.*EEG-fMRI neurofeedback of a motor imagery task*, June 2016, Organization for Human Brain Mapping, Poster, https://hal.inria.fr/hal-01426182.