RESEARCH CENTER
Rennes - Bretagne-Atlantique

FIELD

Activity Report 2016

Section Application Domains

Edition: 2017-08-25

ALGORITHMICS, PROGRAMMING, SOFTWARE AND ARCHITECTURE
1. CAIRN Project-Team 4
2. CELTIQUE Project-Team (section vide) 5
3. HYCOMES Project-Team 6
4. PACAP Project-Team 8
5. SUMO Project-Team 9
6. TAMIS Team 11
7. TASC Project-Team 12
8. TEA Project-Team 13
APPLIED MATHEMATICS, COMPUTATION AND SIMULATION
9. ANJA Team 15
10. ASPI Project-Team 22
11. I4S Project-Team 23
12. IPSO Project-Team (section vide) 25
DIGITAL HEALTH, BIOLOGY AND EARTH
13. DYLISS Project-Team 26
14. FLUMINANCE Project-Team 29
15. GENSCALE Project-Team 30
16. SERPICO Project-Team 31
17. VISAGES Project-Team 33
NETWORKS, SYSTEMS AND SERVICES, DISTRIBUTED COMPUTING
18. ASAP Project-Team (section vide) 34
19. ASCOLA Project-Team 35
20. CIDRE Project-Team 37
21. DIONYSOS Project-Team 38
22. DIVERSE Project-Team 39
23. KERDATA Project-Team 40
24. MYRIADS Project-Team 42
25. TACOMA Team 43
PERCEPTION, COGNITION AND INTERACTION
26. HYBRID Project-Team 45
27. LACODAM Team 46
28. LAGADIC Project-Team 49
29. LINKMEDIA Project-Team 50
30. MIMETIC Project-Team 51
31. PANAMA Project-Team 55
32. SIROCCO Project-Team 56

CAIRN Project-Team

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.

CELTIQUE Project-Team (section vide)

HYCOMES Project-Team

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 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).

[0] http://www.ceser-reunion.fr/fileadmin/user_upload/tx_pubdb/archives/10.10.18_Rapport_electricite.pdf

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]).

[0] https://www.fmi-standard.org/
[0] http://www.adacore.com/gnatpro-safety-critical/avionics/do178c/

PACAP Project-Team

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.

SUMO Project-Team

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.

TAMIS Team

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.

TASC Project-Team

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.

TEA Project-Team

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.

ANJA Team

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.

[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.

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.
- 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.

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:

[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.

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.

ASPI Project-Team

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.

I4S Project-Team

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.

IPSO Project-Team (section vide)

DYLISS Project-Team

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).

FLUMINANCE Project-Team

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.

.

GENSCALE Project-Team

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.

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.

SERPICO Project-Team

4. Application Domains

4.1. Modeling and analysis of membrane transport and molecule trafficking at the single cell scale

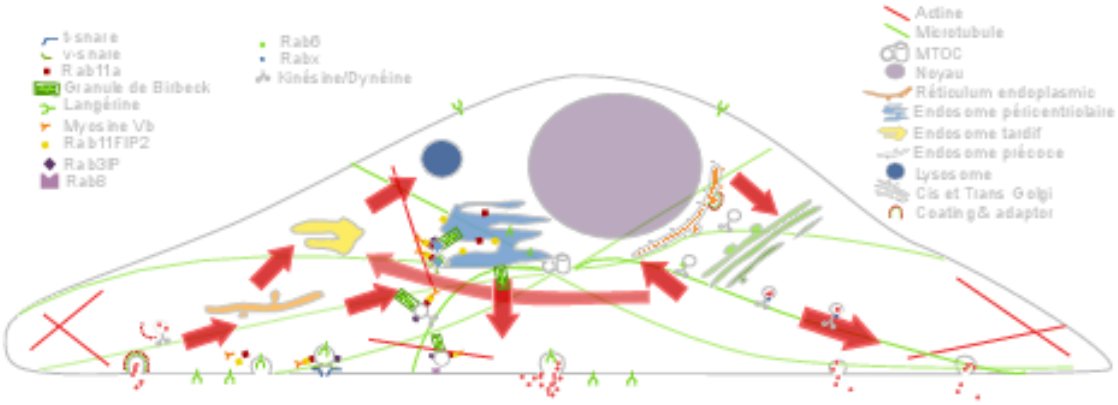

Figure 1. Cargo Langerin Trafficking controlled by Rab11A/Rab11FIP2/MyoVb platform.

In the past recent years, research carried at UMR 144 CNRS-Institut Curie ("Space Time imaging of Endomembranes and organelles Dynamics" team) contributed to a better understanding of the intracellular compartimentation of specialized model cells such as melanocytes and Langerhans cells, the components and structural events involved in the biogenesis of their specialized organelles: melanosomes and Birbeck granules, respectively. These studies have started to highlight: i/ multiple sorting and structural events involved in the biogenesis of these organelles; ii/ complexity of the endo-melanosomal network of these highly specialized cells; iii/ complex molecular architecture organizing and coordinating their dynamics; iv/ intracellular transport steps affected in genetic diseases, among which the Hermansky Pudlak syndrome (HPS) or involved in viral infection (HIV and Langerin in Langerhans cells).

In this context, the central aim of SERPICO is to understand how the different machineries of molecular components involved are interconnected and coordinated to generate such specialized structures. 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.

VISAGES Project-Team

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.

ASAP Project-Team (section vide)

ASCOLA Project-Team

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.

CIDRE Project-Team

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.

DIONYSOS Project-Team

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.

DIVERSE Project-Team

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.

KERDATA Project-Team

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).

MYRIADS Project-Team

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.

TACOMA Team

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**.

HYBRID Project-Team

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.

LACODAM Team

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...).

[0] PEPS: Pharmaco-Epidémiologie et Produits de Santé – Pharmacoepidemiology of health products

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.

LAGADIC Project-Team

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.

LINKMEDIA Project-Team

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.

MIMETIC Project-Team

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.

PANAMA Project-Team

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.

SIROCCO Project-Team

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.