

Inria

RESEARCH CENTER
Rennes - Bretagne-Atlantique

FIELD

Activity Report 2019

Section Application Domains

Edition: 2020-03-21

1. CAIRN Project-Team	4
2. CELTIQUE Project-Team (section vide)	5
3. CIDRE Project-Team (section vide)	6
4. DIONYSOS Project-Team	7
5. DIVERSE Project-Team (section vide)	8
6. DYLISS Project-Team	9
7. EASE Project-Team	10
8. EMPENN Project-Team	12
9. FLUMINANCE Project-Team (section vide)	15
10. GALLINETTE Project-Team (section vide)	16
11. GENSCALE Project-Team	17
12. HYBRID Project-Team	18
13. HYCOMES Project-Team (section vide)	19
14. I4S Project-Team (section vide)	20
15. KERDATA Project-Team	21
16. LACODAM Project-Team	22
17. LINKMEDIA Project-Team	26
18. MIMETIC Project-Team	27
19. MINGUS Project-Team	31
20. Myriads Project-Team	33
21. PACAP Project-Team	34
22. PANAMA Project-Team	35
23. RAINBOW Project-Team (section vide)	36
24. SERPICO Project-Team	37
25. SIMSMART Project-Team	39
26. SIROCCO Project-Team	40
27. STACK Project-Team	42
28. SUMO Project-Team	44
29. TAMIS Project-Team (section vide)	46
30. TEA Project-Team	47
31. WIDE Project-Team (section vide)	49

CAIRN Project-Team

4. Application Domains

4.1. Panorama

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

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

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

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

CELTIQUE Project-Team (section vide)

CIDRE Project-Team (section vide)

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 (section vide)

DYLISS Project-Team

4. Application Domains

4.1. Application fields in biology

In terms of transfer and societal impact, we consider that our role is to develop fruitful collaborations with laboratories of biology in order to consolidate their studies by a smart use of our tools and prototypes and to generate new biological hypotheses to be tested experimentally.

Marine Biology: seaweed enzymes and metabolism & sea-urchin cell-cycle. Our main field of study is **marine biology**, as it is a transversal field covering challenges in integrative biology, dynamical systems and sequence analysis. Our methods based on combinatorial optimization for the reconstruction of genome-scale metabolic networks and on classification of enzyme families based on local and partial alignments allowed the seaweed metabolism *E. Siliculosus* to be deciphered [62], [53]. The study of the *HAD* superfamily of proteins thanks to partial local alignments, produced by *Protomata* tools, allows sub-families to be deciphered and classified, and the metabolic map reconstructed with *Meneco* enabled the reannotation of 56 genes within the *E. siliculosus* genome. These approaches also shed light on evolution of metabolic processes. As a further study, we reconstructed the metabolic network of a symbiot bacterium *Ca. P. ectocarpi* [55] and used this reconstructed network to decipher interactions within the algal-bacteria holobiont, revealing several candidates metabolic pathways for algal-bacterial interactions. Similarly, our analyses suggest that the bacterium *Ca. P. ectocarpi* is able to provide both β -alanine and vitamin B5 to the seaweed via the phosphopantothenate biosynthesis pathway [63].

Micro-biology: elucidating the functioning of extremophile consortiums of bacteria. In this application field, our main issue is the understanding of bacteria living in extreme environments, mainly in collaboration with the group of bioinformatics at Universidad de Chile (co-funded by the Center of Mathematical Modeling, the Center of Regulation Genomics and Inria-Chile). In order to elucidate the main characteristics of these bacteria, our integrative methods were developed to identify the main groups of regulators for their specific response in their living environment. The integrative biology tools *Meneco*, *Lombarde* and *Shogen* have been designed in this context. In particular, genome-scale metabolic network been recently reconstructed and studied with the *Meneco* and *Shogen* approaches, especially on bacteria involved in biomining processes [48] and in Salmon pathogenicity [52].

Agriculture and environmental sciences: upstream controllers of cow, pork and pea-aphid metabolism and regulation. In this application field, our goal is to propose methods to identify regulators of very complex phenotypes related to environmental issues. Our work on the identification of upstream regulators within large-scale knowledge databases (prototype *KeyRegulatorFinder*) [47] and on semantic-based analysis of metabolic networks [45] was very valuable for interpreting differences of gene expression in pork meat [60] and figure out the main gene-regulators of the response of porks to several diets [59]. In addition, constraints-based programming also allows us to decipher regulators of reproduction for the pea aphid, an insect that is a pest on plants. In terms of biological output of the network studies on the pea aphid microRNAs, we have identified one new microRNA (apmir-3019, not present in any known species other than the pea aphid) who has more than 900 putative mRNA targets.

Health: deciphering pathways involved in the TGF- β signalling network. TGF- β is a multifunctional cytokine that regulates mammalian development, differentiation, and homeostasis with both beneficial anti-tumor effect [49] and pro-tumor effect [61]. Deciphering protumor versus antitumor signaling requires to take into account a system-wide view and develop predictive models for therapeutic benefit. For that purpose we developed *Cadbiom* and identified gene networks associated with innate immune response to viral infection that combine TGF- β and interleukine signaling pathways [43], [51].

EASE Project-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*. EASE proposed instead the adoption of a *physical approach*, in which the information is spread in the environment, carried by the entities themselves, and the elaboration is directly executed by these entities "inside" the physical space. This allows performing meaningful exchanges of data that will thereafter need a less complicated processing compared to the current solutions. The result is a smart environment that can, in an easier and better way, integrate the context in its functioning and thus seamlessly deliver more useful and effective user services. Our contribution aims at implementing the physical approach in a smarter environment, showing a solution for improving both comfort and energy savings.

4.2. Automation in Smart City

The domain of Smart Cities is still young but it is already a huge market which attracts number of companies and researchers. It is also multi-fold as the words "smart city" gather multiple meanings. Among them one of the main responsibilities of a city, is to organize the transportation of goods and people. In intelligent transportation systems (ITS), ICT technologies have been involved to improve planification and more generally efficiency of journeys within the city. We are interested in the next step where efficiency would be improved locally relying on local interactions between vehicles, infrastructure and people (smartphones).

For the future autonomous vehicle are now in the spotlight, since a lot of works has been done in recent years in automotive industry as well as in academic research centers. Such unmanned vehicles could strongly impact the organisation of the transportation in our cities. However, due to the lack of a definition of what is an "autonomous" vehicle, it is still difficult to see how these vehicles will interact with their environment (eg. road, smart city, houses, grid, etc.). From augmented perception to fully cooperative automated vehicle, the autonomy covers various realities in terms of interaction the vehicle relies on. The extended perception relies on communication between the vehicle and surrounding roadside equipments. That helps the driving system to build and maintain an accurate view of the environment. But at this first stage the vehicle only uses its own perception to make its decisions. At a second stage, it will take advantages of local interaction with other vehicles through car-to-car communications to elaborate a better view of its environment. Such "cooperative autonomy" does not try to reproduce the human behavior anymore, it strongly relies on communication between vehicles and/or with the infrastructure to make decision and to acquire information on the environment. Part of the decision could be centralized (almost everything for an automatic metro) or coordinated by a roadside component. The decision making could even be fully distributed but this put high constraints on the communications. Automated vehicles are just an example of smart city automated processes that will have to share information within the surrounding to make their decisions.

4.3. Pervasive applications in uncontrolled environments

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 as 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 principles 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 a 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 a large number of objects of various nature, arranged randomly in a given area or container. **Most pervasive computing applications fall in this context.**

EMPENN Project-Team

4. Application Domains

4.1. Population imaging

One major objective of neuroimaging researchers and clinicians is to be able to stratify brain imaging data in order to derive new and more specific population models. In practice this requires to set up large-scale experiments that, due to the lack of resources and capabilities to recruit locally subjects who meet specific inclusion criteria, motivates the need for sharing the load.

But, building and using multi-site large-scale resources poses specific challenges to deal with the huge quantity of data produced and their diversity. Empenn will focus on two challenges in particular:

- Provide computational environments for the computation and use of imaging biomarkers in the targeted brain diseases, a solution to be used by radiologists and neurologists/psychiatrists for the clinical follow-up of a large patient population.
- Modeling analytic variability of image processing pipelines to better understand and predict the behaviour of imaging biomarker detection solutions and improve reproducibility and productivity in clinical neuroimaging research.

4.2. Detection and learning

We intend to make significant contributions with major impacts in learning coupling models between functional recordings during neurofeedback procedures. These advances will provide a breakthrough in brain-computer interfaces for rehabilitation protocols. Our aim is to:

- Provide a computational environment that combines data-driven (machine learning) and Bayesian solutions to improve the detection of abnormal patterns in images through decision or evidence theory data fusion strategies. The major initial application will be for multiple sclerosis. Over the longer term, we also expect to adapt these methods to address a wider range of neurological diseases (epilepsy, stroke, tumors, etc.) in neonate and adult brains.
- Develop solutions for combining brain state measurements from multimodal sensors or sequences (e.g. fMRI, ASL, EEG, NIRS, etc.) with applications in the spatiotemporal reconstruction of brain activity from MRI-EEG or the combined detection of the endogenous hemodynamic and resting state network of the brain from ASL and NIRS. Over the longer term, the advent of new hybrid brain imaging sensors (e.g. PET-MRI) will require these methods to be extended to a larger spectrum of information combining structural, morphological, metabolic, electrophysiological and cellular/molecular information (e.g. through the use of specific ligands/nanocarriers).

4.3. Quantitative imaging

The Empenn research group focuses on the development of several quantitative techniques in magnetic resonance imaging of the brain. These methods allow for a characterization of both the function and the structure of the brain with high precision. Arterial spin labelling (ASL) is a contrast agent-free imaging technique which labels arterial blood water as an endogenous tracer for perfusion and can measure resting-state cerebral blood flow. We are interested in estimating multiparametric hemodynamics using ASL, such as combined cerebral blood flow and arterial transit times, and derive statistical descriptors to represent significant differences between groups. In addition to quantitative perfusion parameters, our contributions on tissue compartment imaging aim at delineating neural circuits and characterize their microstructure properties, using both diffusion MRI and relaxometry. In diffusion MRI, arbitrary gradient waveforms were shown to exhibit higher sensitivity to microstructure parameters than standard pulsed gradients. We work on the optimization of sampling protocols in this domain, with the objective to propose sequences compatible with in vivo acquisition.

Complementary to diffusion MRI, we develop methods for the reconstruction of myelin-bound, extra-axonal and cerebrospinal fluid water using multi-compartment modelling of the T2-relaxometry signal. We combine these techniques with tractography to identify trajectories of pathologies associated to the evolution of these microstructural parameters along specific fiber bundles in the brain white matter.

4.4. Behavior

Advances in the field of in vivo imaging offer new opportunities for addressing the management of resistant affective disorders and their consequences (suicide risk and socio-professional impact), and the management of spatial cognition disorders after stroke and their consequences (postural perturbations and the loss of autonomy). Our objective, and the main challenge in this context, will be to introduce medical image computing methods to the multidisciplinary field of behavioral disorders (cognitive disorders, particularly spatial and postural control disorders or anterograde memory impairment, mood disorders, notably resistant depression, schizophrenic disorders, pervasive developmental disorders, attention disorders, etc.) in order to gain a better understanding of the pathology and devise innovative therapeutic approaches.

We also expect to become a major player in the future and make important contributions with significant impacts, primarily in drug-resistant depression in young and old populations. In particular, we expect to provide new image-related metrics combining perfusion, metabolism and microstructural information regarding the brain in order to better characterize pathologies, provide prospective evolution values and potentially provide new brain stimulation targets that could be used in neurofeedback rehabilitation protocols or other types of brain stimulation procedure.

We aim to provide new imaging markers of mental diseases, especially in the context of mood disorders. The new biomarkers will be derived from the metabolic (ASL and later ASL+PET) point of view as well as from the microstructural point of view (multicompartment diffusion MRI and relaxometry). Similarly, we expect to exhibit imaging biomarker regularities combining metabolic and structural information. Over the longer term, we expect these biomarkers to be the target of neurofeedback rehabilitation procedures. Also, over the longer term, we expect to supplement the MRI markers with molecular marker ones coming from new PET tracers, especially those associated with serotonin intake, at one time point or during a rehabilitation protocol under hybrid PET-EEG-MRI neurofeedback procedures.

4.5. Neuroinflammation

Some of the major ongoing research issues in the neuroimaging of neuro-inflammatory diseases concern the definition of new biomarkers for tracking the development of the pathology using high-dimensional data (e.g. nD+t MRI). This includes the use of white matter-specific imaging, such as magnetization transfer MRI, relaxometry and diffusion-weighted imaging (DW-MRI). Our objective is (1) to develop information-processing tools to tag the spatiotemporal evolutions of MS patterns at the brain parenchyma and spinal cord levels from their different signatures (inflammatory cells visible with USPIO or Gd contrast agents on MRI, persistent black holes, eloquent regional atrophy and microstructure signatures); and (2) to test these new tools on new imaging cohorts. In this respect, we for instance conduct studies on brain and spinal cord imaging, continuing on from the PHRC multicentric EMISEP project (PI: G. Edan), as it is very likely that lesions in the spine will directly affect the ambulatory ability of the patient (and thereby the clinical scores). In order to extend this experiment to a larger MS population, based on our expertise from the OFSEP cohort, we also plan to improve the MS therapeutic decision process through the MUSIC project (Multiple Sclerosis Imaging Check out, a public/private project). Our goal is to develop and assess a standardized monitoring tool that provides a robust, long-term computerized MRI follow-up that will become the gold standard in clinical practice for therapeutic decisions in MS treatment. As part of this project, Empenn will share its expertise in data management systems (Shanoir and FLI-IAM) and automatic processing tools (through the medInria and Anima software repositories) to extract quantitative indices from the images.

4.6. Recovery

Mental and neurological disorders are the leading cause of years lived with a disability. Treatment-resistant depression affects approximately 2% of the European population. Meanwhile, in the case of brain disorders, almost 1.5 million Europeans (15 million people worldwide) suffer a stroke event each year. Current recovery methods for brain disorders and traumatic brain injuries remain limited, preventing many from achieving full recuperation. We propose addressing the issue of brain recovery by introducing new advances from recent breakthroughs in computational medical imaging, data processing and human-machine interfaces, and demonstrating how these new concepts can be used, in particular for the treatment of stroke and major depressive disorders.

We ambition to combine advanced instrumental devices (Hybrid EEG, NIRS and MRI platforms), with new hybrid brain computer interface paradigms and new computational models to provide neurofeedback-based therapeutic and neuro-rehabilitation paradigms in some of the major mental and neurological disorders of the developmental and the aging brain.

Neurofeedback involves using a brain-computer interface that provides an individual with real-time biofeedback about his or her brain activity in the form of sensory feedback. It enables individuals to learn to better control their brain activity, which can be measured in real time using various non-invasive sensors as described above. Although EEG is currently the only modality used by clinical practitioners in that context, it lacks specificity due to its low spatial resolution. Dynamic research into fMRI-neurofeedback has held promise for treating depression, chronic pain and stroke, since it offers the prospect of real-time imagery of the activity in deep brain structures with high spatial resolution. However, the low temporal resolution and high cost of fMRI-Neurofeedback has hampered the development of many applications. We believe that the future belongs to hybrid responses that combine multimodal sensors and intend to demonstrate this in the Empenn project.

FLUMINANCE Project-Team (section vide)

GALLINETTE Project-Team (section vide)

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

Genetic and cancer disease diagnostic: Genetic diseases are caused by some particular mutations in the genomes that alter important cell processes. Similarly, cancer comes from changes in the DNA molecules that alter cell behavior, causing uncontrollable growth and malignancy. Pointing out genes with mutations helps in identifying the disease and in prescribing the right drug. Thus, DNA from individual patients is sequenced and the aim is to detect potential mutations that may be linked to the patient disease. Bioinformatics analysis can be based on the detection of SNPs (Single Nucleotide Polymorphism) from a set of predefined target genes. One can also scan the complete genome and report all kinds of mutations, including complex mutations such as large insertions or deletions, that could be associated with genetic or cancer diseases.

Neurodegenerative disorders: The biological processes that lead from abnormal protein accumulation to neuronal loss and cognitive dysfunction is not fully understood. In this context, neuroimaging biomarkers and statistical methods to study large datasets play a pivotal role to better understand the pathophysiology of neurodegenerative disorders. The discovery of new genetic biomarkers could thus have a major impact on clinical trials by allowing inclusion of patients at a very early stage, at which treatments are the most likely to be effective. Correlations with genetic variables can determine subgroups of patients with common anatomical and genetic characteristics.

4.3. Agronomy

Insect genomics: Insects represent major crop pests, justifying the need for control strategies to limit population outbreaks and the dissemination of plant viruses they frequently transmit. Several issues are investigated through the analysis and comparison of their genomes: understanding their phenotypic plasticity such as their reproduction mode changes, identifying the genomic sources of adaptation to their host plant and of ecological speciation, and understanding the relationships with their bacterial symbiotic communities [6].

Improving plant breeding: Such projects aim at identifying favorable alleles at loci contributing to phenotypic variation, characterizing polymorphism at the functional level and providing robust multi-locus SNP-based predictors of the breeding value of agronomical traits under polygenic control. Underlying bioinformatics processing is the detection of informative zones (QTL) on the plant genomes.

4.4. Environment

Food quality control: One way to check food contaminated with bacteria is to extract DNA from a product and identify the different strains it contains. This can now be done quickly with low-cost sequencing technologies such as the MinION sequencer from Oxford Nanopore Technologies.

Ocean biodiversity: The metagenomic analysis of seawater samples provides an original way to study the ecosystems of the oceans. Through the biodiversity analysis of different ocean spots, many biological questions can be addressed, such as the plankton biodiversity and its role, for example, in the CO₂ sequestration.

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 movie industry, content customization, video games or attractions in theme parks,
- **Construction:** with virtual mock-ups design and review, or historical/architectural visits.
- **Cultural Heritage:** with acquisition, virtual excavation, virtual reconstruction and visualization

HYCOMES Project-Team (section vide)

I4S Project-Team (section vide)

KERDATA Project-Team

4. Application Domains

4.1. Application Domains

The KerData team investigates the design and implementation of architectures for data storage and processing across clouds, HPC and edge-based systems, which address the needs of a large spectrum of applications. The use cases we target to validate our research results come from the following domains.

- Climate and meteorology
- Earth science
- Energy and sustainable development
- Smart cities
- Data science

LACODAM Project-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 not the exception. We are eager to see our work applied in real world applications, and have thus an important activity in maintaining strong ties with industrial partners concerned with marketing and energy as well as public partners working on health, agriculture and environment.

4.2. Industry

We present below our industrial collaborations. Some are well established partnerships, while others are more recent collaborations with local industries that wish to reinforce their Data Science R&D with us (e.g. Energiency, Amossys).

- **Resource Consumption Analysis for Optimizing Energy Consumption and Practices in Industrial Factories (Energiency).** In order to increase their understanding of factory operation, companies introduce more and more sensors in their factories. Thus, the resource (electricity, water, etc.) consumption of engines, workshops and factories are recorded in the form of times series or temporal sequences. The person who is in charge of resource consumption optimization needs better software than classical spreadsheets for this purpose. He/she needs effective decision-aiding tools with statistical and artificial intelligence knowledge. The start-up Energiency aims at designing and offering such pieces of software for analyzing energy consumption. The CIFRE PhD thesis of Maël (defended 16/12/2019) aimed at proposing new approaches and solutions from the data mining field to tackle this issue.
- **Security (Amossys).** Current networks are faced with an increasing variety of attacks, from the classic “DDoS” that makes a server unusable for a few hours, to advanced attacks that silently infiltrate a network and exfiltrate sensitive information months or even years later. Such intrusions, called APT (Advanced Persistent Threat) are extremely hard to detect, and this will become even harder as most communications will be encrypted. A promising solution is to work on “behavioral analysis”, by discovering patterns based on the metadata of IP-packets. Such patterns can relate to an unusual sequencing of events, or to an unusual communication graph. Finding such complex patterns over a large volume of streaming data requires to revisit existing stream mining algorithms to dramatically improve their throughput, while guaranteeing a manageable false positive rate. We collaborated on this topic with the Amossys company and the EMSEC team of Irista through the co-supervision of the CIFRE PhD of Alban Siffer (located in the EMSEC team, defended 19/12/2019). Our goal was in particular to design novel anomaly detection methods making minimal assumptions on the data, and able to scale to real traffic volumes.
- **Car Sharing Data Analysis.** Peugeot-Citroën (PSA) group’s know-how encompasses all areas of the automotive industry, from production to distribution and services. Among others, its aim is to provide a car sharing service in many large cities. This service consists in providing a fleet of cars and a “free floating” system that allows users to use a vehicle, then drop it off at their convenience in the city. To optimize their fleet and the availability of the cars throughout the city, PSA needs to analyze the trajectory of the cars and understand the mobility needs and behavior of their users. We tackle this subject together through the CIFRE PhD of Gregory Martin.
- **Multimodal Data Analysis for the Supervision of Sensitive Sites.** ATERMES is an international mid-sized company with a strong expertise in high technology and system integration from the upstream design to the long-life maintenance cycle. It has recently developed a new product, called BARRIER TM (“Beacon Autonomous Reconnaissance Identification and Evaluation Response”),

which provides operational and tactical solutions for mastering borders and areas. Once in place, the system allows for a continuous night and day surveillance mission with a small crew in the most unexpected rugged terrain. The CIFRE PhD of Heng Zhang aims at developing a deep learning architecture and algorithms able to detect anomalies (mainly persons) from multimodal data. The data are “multimodal” because information about the same phenomenon can be acquired from different types of detectors, at different conditions, in multiple experiments.

- **Root Cause Analysis in Networks.** AdvisorSLA is a French company specialized in software solutions for network monitoring. For this purpose, the company relies on techniques of network metrology. By continuously measuring the state of the network, monitoring solutions detect events (e.g., overloaded router) that may degrade the network’s operation and the quality of the services running on top of it (e.g., video transmission could become choppy). When a monitoring solution detects a potentially problematic sequence of events, it triggers an alarm so that the network manager can take actions. Those actions can be preventive or corrective. Some statistics show that only 40% of the triggered alarms are conclusive, that is, they manage to signal a well-understood problem that requires an action from the network manager. This means that the remaining 60% are presumably false alarms. While false alarms do not hinder network operation, they do incur an important cost in terms of human resources. Thus, the CIFRE PhD of Erwan Bourrand proposes to characterize conclusive and false alarms. This will be achieved by designing automatic methods to “learn” the conditions that most likely precede each type of alarm, and therefore predict whether the alarm will be conclusive or not. This can help adjust existing monitoring solutions in order to improve their accuracy. Besides, it can help network managers automatically trace the causes of a problem in the network.

4.3. Health

- **Care Pathways Analysis for Supporting Pharmaco-Epidemiological Studies.** Pharmaco-epidemiology applies the methodologies developed in general epidemiology to answer to questions about the uses and effects of health products, drugs [31], [30] or medical devices [23], on population. In classical pharmaco-epidemiology studies, people who share common characteristics are recruited to build a dedicated prospective cohort. Then, meaningful data (drug exposures, diseases, etc.) are collected from the cohort within a defined period of time. Finally, a statistical analysis highlights the links (or the lack of links) between drug exposures and outcomes (e.g., adverse effects). The main drawback of prospective cohort studies is the time required to collect the data and to integrate them. Indeed, in some cases of health product safety, health authorities have to answer quickly to pharmaco-epidemiology questions.

New approaches of pharmaco-epidemiology consist in using large EHR (Electronic Health Records) databases to investigate the effects and uses (or misuses) of drugs in real conditions. The objective is to benefit from nationwide available data to answer accurately and in a short time pharmaco-epidemiological queries for national public health institutions. Despite the potential availability of the data, their size and complexity make their analysis long and tremendous. The challenge we tackle is the conception of a generic digital toolbox to support the efficient design of a broad range of pharmaco-epidemiology studies from EHR databases. We propose to use pattern mining algorithms and reasoning techniques to analyse the typical care pathways of specific groups of patients.

To answer the broad range of pharmaco-epidemiological queries from national public health institutions, the PEPS⁰ platform exploits, in secondary use, the French health cross-schemes insurance system, called SNDS. The SNDS covers most of the French population with a sliding period of 3 past years. The main characteristics of this data warehouse are described in [29]. Contrary to local hospital EHR or even to other national initiatives, the SNDS data warehouse covers a huge population. It makes possible studies on unfrequent drugs or diseases in real conditions of use. To tackle the volume and the diversity of the SNDS data warehouse, a research program has been established

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

to design an innovative toolbox. This research program is focused first on the modeling of care pathways from the SNDS database and, second, on the design of tools supporting the extraction of insights about massive and complex care pathways by clinicians. In such a database a care pathway is an individual sequence of drugs exposures, medical procedures and hospitalizations.

- **Care Sequences for the Exploration of Medico-administrative Data.** The difficulty of analyzing medico-administrative data is the semantic gap between the raw data (for example, database record about the delivery at date t of drug with ATC 2 code N 02BE01) and the nature of the events sought by clinicians (“was the patient exposed to a daily dose of paracetamol higher than 3g?”). The solution that is used by epidemiologists consists in enriching the data with new types of events that, on the one side, could be generated from raw data and on the other side, have a medical interpretation. Such new abstract events are defined by clinician using proxies. For example, drugs deliveries can be translated in periods of drug exposure (drug exposure is a time-dependent variable for non-random reasons) or identify patient stages of illness, etc. A proxy can be seen as an abstract description of a care sequence.

Currently, the clinicians are limited in the expression of these proxies both by the coarse expressivity of their tools and by the need to process efficiently large amount of data. From a semantic point of view, care sequences must fully integrate the temporal and taxonomic dimensions of the data to provide significant expression power. From a computational point of view, the methods employed must make it possible to efficiently handle large amounts of data (several millions care pathways). The aim of the PhD of Johanne Bakalara is to study temporal models of sequences in order 1) to show their abilities to specify complex proxies representing care sequences needed in pharmaco-epidemiological studies and 2) to build an efficient querying tool able to exploit large amount of care pathways.

4.4. Agriculture and environment

- **Dairy Farming.** The use and analysis of data acquired in dairy farming is a challenge both for data science and animal science. The goal is to improve farming conditions, i.e., health, welfare and environment, as well as farmers’ income. Nowadays, animals are monitored by multiple sensors giving a wealth of heterogeneous data such as temperature, weight, or milk composition. Current techniques used by animal scientists focus mostly on mono-sensor approaches. The dynamic combination of several sensors could provide new services and information useful for dairy farming. The PhD thesis of Kevin Fauvel (#DigitAg grant), aims to study such combinations of sensors and to investigate the use data mining methods, especially pattern mining algorithms. The challenge is to design new algorithms that take into account data heterogeneity—in terms of nature and time units—, and that produce useful patterns for dairy farming. The outcome of this thesis will be an original and important contribution to the new challenge of the IoT (Internet of Things) and will interest domain actors to find new added value to a global data analysis. The PhD thesis, started on October 2017, takes place in an interdisciplinary setting bringing together computer scientists from Inria and animal scientists from INRA, both located in Rennes.
- **Optimizing the Nutrition of Individual Sow.** Another direction for further research is the combination of data flows with prediction models in order to learn nutrition strategies. Raphaël Gauthier started a PhD thesis (#DigitAg Grant) in November 2017 with both Inria and INRA supervisors. His research addresses the problem of finding the optimal diet to be supplied to individual sows. Given all the information available, e.g., time-series information about previous feeding, environmental data, scientists models, the research goal is to design new algorithms to determine the optimal ration for a given sow in a given day. Efficiency issues of developed algorithms will be considered since the proposed software should work in real-time on the automated feeder. The decision support process should involve the stakeholder to ensure a good level of acceptance, confidence and understanding of the final tool.
- **Ecosystem Modeling and Management.** Ongoing research on ecosystem management includes

modelling of ecosystems and anthropogenic pressures, with a special concern on the representation of socio-economical factors that impact human decisions. A main research issue is how to represent these factors and how to integrate their impact on the ecosystem simulation model. This work is an ongoing cooperation with ecologists from the Marine Spatial Ecology of Queensland University, Australia and from Agrocampus Ouest.

4.5. Education

- **Data-oriented Academic Counseling.** Course selection and recommendation are important aspects of any academic counseling system. The Learning Analytics community has long supported these activities via automatic, data-based tools for recommendation and prediction. LACODAM, in collaboration with the Ecuadorian research center CTI⁰ has contributed to this body of research with the design of a tool that allows students to select multiple courses and predict their academic performance based on historical academic data. The tool resorts to interpretable machine learning techniques, and is intended to be used by the students before the counseling sessions to plan their upcoming semester at the Ecuadorian university ESPOL. The data visualization aspects of the tool, as well as the data science considerations and our emphasis on explainability are compiled in a paper that is under revision at the Learning Analytics and Knowledge Conference (LAK'20). The data science component of the tool was developed during the M1 internship of Mohammad Poul-Doust.

4.6. Others

- **RDF Archiving.** The dynamicity of the Semantic Web has motivated the development of solutions for RDF archiving, i.e., the task of storing and querying all previous versions of an RDF dataset. Notwithstanding the value of RDF archiving for data maintainers and consumers, this field of research remains under-developed for multiple reasons. These include notably (i) the lack of usability and scalability of the existing systems, (ii) no archiving support for multi-graph RDF datasets, (iii) the absence of a standard SPARQL extension for RDF archives, and (iv) a disregard of the evolution patterns of RDF datasets. The PhD thesis of Olivier Pelgrin aims at developing techniques towards a scalable and full-fledged archiving solution for the Semantic Web. This PhD thesis is a collaboration between LACODAM and the DAISY team at Aalborg University.

⁰Centro de Tecnologías de Información, <http://cti.espol.edu.ec/>

LINKMEDIA Project-Team

4. Application Domains

4.1. Asset management in the entertainment business

Media asset management—archiving, describing and retrieving multimedia content—has turned into a key factor and a huge business for content and service providers. Most content providers, with television channels at the forefront, rely on multimedia asset management systems to annotate, describe, archive and search for content. So do archivists such as the Institut National de l’Audiovisuel, the bibliothèque Nationale de France, the Nederlands Instituut voor Beeld en Geluid or the British Broadcast Corporation, as well as media monitoring companies, such as Yacast in France. Protecting copyrighted content is another aspect of media asset management.

4.2. Multimedia Internet

One of the most visible application domains of linked multimedia content is that of multimedia portals on the Internet. Search engines now offer many features for image and video search. Video sharing sites also feature search engines as well as recommendation capabilities. All news sites provide multimedia content with links between related items. News sites also implement content aggregation, enriching proprietary content with user-generated content and reactions from social networks. Most public search engines and Internet service providers offer news aggregation portals. This also concerns TV on-demand and replay services as well as social TV services and multi-screen applications. Enriching multimedia content, with explicit links targeting either multimedia material or knowledge databases is central here.

4.3. Data journalism

Data journalism forms an application domain where most of the technology developed by LINKMEDIA can be used. On the one hand, data journalists often need to inspect multiple heterogeneous information sources, some being well structured, some other being fully unstructured. They need to access (possibly their own) archives with either searching or navigational means. To gradually construct insight, they need collaborative multimedia analytics processes as well as elements of trust in the information they use as foundations for their investigations. Trust in the information, watching for adversarial and/or (deep) fake material, accountability are all crucial here.

MIMETIC Project-Team

4. Application Domains

4.1. Animation, Autonomous Characters and Digital Storytelling

Computer Animation is one of the main application domain of the research work conducted in the MimeTIC team, in particular in relation to the entertainment and game industries. In these domains, creating virtual characters that are able to replicate real human motions and behaviours still highlights unanswered key challenges, especially as virtual characters are becoming more and more required to populate virtual worlds. For instance, virtual characters are used to replace secondary actors and generate highly populated scenes that would be hard and costly to produce with real actors, which requires to create high quality replicas that appear, move and behave both individually and collectively like real humans. The three key challenges for the MimeTIC team are therefore (i) to create natural animations (i.e., virtual characters that move like real humans), (ii) to create autonomous characters (i.e., that behave like real humans) and (iii) to orchestrate the virtual characters so as to create interactive stories.

First, our challenge is therefore to create animations of virtual characters that are natural, in the largest sense of the term of moving like a real human real would. This challenge covers several aspects of Character Animation depending on the context of application, e.g., producing visually plausible or physically correct motions, producing natural motion sequences, etc. Our goal is therefore to develop novel methods for animating virtual characters, e.g., based on motion capture, data-driven approaches, or learning approaches. However, because of the complexity of human motion (e.g., the number of degrees of freedom that can be controlled), resulting animations are not necessarily physically, biomechanically, or visually plausible. For instance, current physics-based approaches produce physically correct motions but not necessarily perceptually plausible ones. All these reasons are why most entertainment industries still mainly rely on manual animation, e.g., in games and movies. Therefore, research in MimeTIC on character animation is also conducted with the goal of validating objective (e.g., physical, biomechanical) as well as subjective (e.g., visual plausibility) criteria.

Second, one of the main challenges in terms of autonomous characters is to provide a unified architecture for the modeling of their behavior. This architecture includes perception, action and decisional parts. This decisional part needs to mix different kinds of models, acting at different time scale and working with different nature of data, ranging from numerical (motion control, reactive behaviors) to symbolic (goal oriented behaviors, reasoning about actions and changes). For instance, autonomous characters play the role of actors that are driven by a scenario in video games and virtual storytelling. Their autonomy allows them to react to unpredictable user interactions and adapt their behavior accordingly. In the field of simulation, autonomous characters are used to simulate the behavior of humans in different kind of situations. They enable to study new situations and their possible outcomes. In the MimeTIC team, our focus is therefore not to reproduce the human intelligence but to propose an architecture making it possible to model credible behaviors of anthropomorphic virtual actors evolving/moving in real time in virtual worlds. The latter can represent particular situations studied by psychologists of the behavior or to correspond to an imaginary universe described by a scenario writer. The proposed architecture should mimic all the human intellectual and physical functions.

Finally, interactive digital storytelling, including novel forms of edutainment and serious games, provides access to social and human themes through stories which can take various forms and contains opportunities for massively enhancing the possibilities of interactive entertainment, computer games and digital applications. It provides chances for redefining the experience of narrative through interactive simulations of computer-generated story worlds and opens many challenging questions at the overlap between computational narratives, autonomous behaviours, interactive control, content generation and authoring tools. Of particular interest for the MimeTIC research team, virtual storytelling triggers challenging opportunities in providing effective models for enforcing autonomous behaviours for characters in complex 3D environments. Offering both low-level capacities to characters such as perceiving the environments, interacting with the environment and

reacting to changes in the topology, on which to build higher-levels such as modelling abstract representations for efficient reasoning, planning paths and activities, modelling cognitive states and behaviours requires the provision of expressive, multi-level and efficient computational models. Furthermore virtual storytelling requires the seamless control of the balance between the autonomy of characters and the unfolding of the story through the narrative discourse. Virtual storytelling also raises challenging questions on the conveyance of a narrative through interactive or automated control of the cinematography (how to stage the characters, the lights and the cameras). For example, estimating visibility of key subjects, or performing motion planning for cameras and lights are central issues for which have not received satisfactory answers in the literature.

4.2. Fidelity of Virtual Reality

VR is a powerful tool for perception-action experiments. VR-based experimental platforms allow exposing a population to fully controlled stimuli that can be repeated from trial to trial with high accuracy. Factors can be isolated and objects manipulations (position, size, orientation, appearance, ..) are easy to perform. Stimuli can be interactive and adapted to participants' responses. Such interesting features allow researchers to use VR to perform experiments in sports, motion control, perceptual control laws, spatial cognition as well as person-person interactions. However, the interaction loop between users and their environment differs in virtual conditions in comparison with real conditions. When a user interact in an environment, movement from action and perception are closely related. While moving, the perceptual system (vision, proprioception,..) provides feedback about the users' own motion and information about the surrounding environment. That allows the user to adapt his/her trajectory to sudden changes in the environment and generate a safe and efficient motion. In virtual conditions, the interaction loop is more complex because it involves several material aspects.

First, the virtual environment is perceived through a numerical display which could affect the available information and thus could potentially introduce a bias. For example, studies observed a distance compression effect in VR, partially explained by the use of Head Mounted Display with reduced field of view and exerting a weight and torques on the user's head. Similarly, the perceived velocity in a VR environment differs from the real world velocity, introducing an additional bias. Other factors, such as the image contrast, delays in the displayed motion and the point of view can also influence efficiency in VR. The second point concerns the user's motion in the virtual world. The user can actually move if the virtual room is big enough or if wearing a head mounted display. Even with a real motion, authors showed that walking speed is decreased, personal space size is modified and navigation in VR is performed with increased gait instability. Although natural locomotion is certainly the most ecological approach, the physical limited size of VR setups prevents from using it most of the time. Locomotion interfaces are therefore required. Locomotion interfaces are made up of two components, a locomotion metaphor (device) and a transfer function (software), that can also introduce bias in the generated motion. Indeed, the actuating movement of the locomotion metaphor can significantly differ from real walking and the simulated motion depends on the transfer function applied. Locomotion interfaces cannot usually preserve all the sensory channels involved in locomotion.

When studying human behavior in VR, the aforementioned factors in the interaction loop potentially introduce bias both in the perception and in the generation of motor behavior trajectories. MimeTIC is working on the mandatory step of VR validation to make it usable for capturing and analyzing human motion.

4.3. Motion Sensing of Human Activity

Recording human activity is a key point of many applications and fundamental works. Numerous sensors and systems have been proposed to measure positions, angles or accelerations of the user's body parts. Whatever the system is, one of the main problems is to be able to automatically recognize and analyze the user's performance according to poor and noisy signals. Human activity and motion are subject to variability: intra-variability due to space and time variations of a given motion, but also inter-variability due to different styles and anthropometric dimensions. MimeTIC has addressed the above problems in two main directions.

Firstly, we have studied how to recognize and quantify motions performed by a user when using accurate systems such as Vicon (product of Oxford Metrics), Qualisys, or Optitrack (product of Natural Point) motion capture systems. These systems provide large vectors of accurate information. Due to the size of the state vector (all the degrees of freedom) the challenge is to find the compact information (named features) that enables the automatic system to recognize the performance of the user. Whatever the method used, finding these relevant features that are not sensitive to intra-individual and inter-individual variability is a challenge. Some researchers have proposed to manually edit these features (such as a Boolean value stating if the arm is moving forward or backward) so that the expertise of the designer is directly linked with the success ratio. Many proposals for generic features have been proposed, such as using Laban notation which was introduced to encode dancing motions. Other approaches tend to use machine learning to automatically extract these features. However most of the proposed approaches were used to seek a database for motions which properties correspond to the features of the user's performance (named motion retrieval approaches). This does not ensure the retrieval of the exact performance of the user but a set of motions with similar properties.

Secondly, we wish to find alternatives to the above approach which is based on analyzing accurate and complete knowledge on joint angles and positions. Hence new sensors, such as depth-cameras (Kinect, product of Microsoft) provide us with very noisy joint information but also with the surface of the user. Classical approaches would try to fit a skeleton into the surface in order to compute joint angles which, again, lead to large state vectors. An alternative would be to extract relevant information directly from the raw data, such as the surface provided by depth cameras. The key problem is that the nature of these data may be very different from classical representation of human performance. In MimeTIC, we try to address this problem in specific application domains that require picking specific information, such as gait asymmetry or regularity for clinical analysis of human walking.

4.4. Sports

Sport is characterized by complex displacements and motions. One main objective is to understand the determinants of performance through the analysis of the motion itself. In the team, different sports have been studied such as the tennis serve, where the goal was to understand the contribution of each segment of the body in the performance but also the risk of injuries as well as other situation in cycling, swimming, fencing or soccer. Sports motions are dependent on visual information that the athlete can pick up in his environment, including the opponent's actions. Perception is thus fundamental to the performance. Indeed, a sportive action, as unique, complex and often limited in time, requires a selective gathering of information. This perception is often seen as a prerogative for action, it then takes the role of a passive collector of information. However, as mentioned by Gibson in 1979, the perception-action relationship should not be considered sequentially but rather as a coupling: we perceive to act but we must act to perceive. There would thus be laws of coupling between the informational variables available in the environment and the motor responses of a subject. In other words, athletes have the ability to directly perceive the opportunities of action directly from the environment. Whichever school of thought considered, VR offers new perspectives to address these concepts by complementary using real time motion capture of the immersed athlete.

In addition to better understand sports and interactions between athletes, VR can also be used as a training environment as it can provide complementary tools to coaches. It is indeed possible to add visual or auditory information to better train an athlete. The knowledge found in perceptual experiments can be for example used to highlight the body parts that are important to look at to correctly anticipate the opponent's action.

4.5. Ergonomics

The design of workstations nowadays tends to include assessment steps in a Virtual Environment (VE) to evaluate ergonomic features. This approach is more cost-effective and convenient since working directly on the Digital Mock-Up (DMU) in a VE is preferable to constructing a real physical mock-up in a Real Environment (RE). This is substantiated by the fact that a Virtual Reality (VR) set-up can be easily modified, enabling quick adjustments of the workstation design. Indeed, the aim of integrating ergonomics evaluation tools in VEs is to facilitate the design process, enhance the design efficiency, and reduce the costs.

The development of such platforms asks for several improvements in the field of motion analysis and VR. First, interactions have to be as natural as possible to properly mimic the motions performed in real environments. Second, the fidelity of the simulator also needs to be correctly evaluated. Finally, motion analysis tools have to be able to provide in real-time biomechanics quantities usable by ergonomists to analyse and improve the working conditions.

In real working condition, motion analysis and musculoskeletal risks assessment raise also many scientific and technological challenges. Similarly to virtual reality, fidelity of the working process may be affected by the measurement method. Wearing sensors or skin markers, together with the need of frequently calibrating the assessment system may change the way workers perform the tasks. Whatever the measurement is, classical ergonomic assessments generally address one specific parameter, such as posture, or force, or repetitions..., which makes it difficult to design a musculoskeletal risk factor that actually represent this risk. Another key scientific challenge is then to design new indicators that better capture the risk of musculoskeletal disorders. However, this indicator has to deal with the tradeoff between accurate biomechanical assessment and the difficulty to get reliable and the required information in real working conditions.

4.6. Locomotion and Interactions between walkers

Modeling and simulating locomotion and interactions between walkers is a very active, complex and competitive domain, interesting various disciplines such as mathematics, cognitive sciences, physics, computer graphics, rehabilitation etc. Locomotion and interactions between walkers are by definition at the very core of our society since they represent the basic synergies of our daily life. When walking in the street, we should produce a locomotor movement while taking information about our surrounding environment in order to interact with people, move without collision, alone or in a group, intercept, meet or escape to somebody. MimeTIC is an international key contributor in the domain of understanding and simulating locomotion and interactions between walkers. By combining an approach based on Human Movement Sciences and Computer Sciences, the team focuses on locomotor invariants which characterize the generation of locomotor trajectories, conducts challenging experiments focusing on visuo-motor coordination involved during interactions between walkers both using real and virtual set-ups. One main challenge is to consider and model not only the "average" behaviour of healthy younger adult but also extend to specific populations considering the effect of pathology or the effect of age (kids, older adults). As a first example, when patients cannot walk efficiently, in particular those suffering from central nervous system affections, it becomes very useful for practitioners to benefit from an objective evaluation of their capacities. To facilitate such evaluations, we have developed two complementary indices, one based on kinematics and the other one on muscle activations. One major point of our research is that such indices are usually only developed for children whereas adults with these affections are much more numerous. We extend this objective evaluation by using person-person interaction paradigm which allows studying visuo-motor strategies deficit in these specific populations.

Another fundamental question is the adaptation of the walking pattern according to anatomical constraints, such as pathologies in orthopedics, or adaptation to various human and non-human primates in paleoanthropology. Hence, the question is to predict plausible locomotion according to a given morphology. This question raises fundamental questions about the variables that are regulated to control gait: balance control, minimum energy, minimum jerk...In MimeTIC we develop models and simulators to efficiently test hypothesis on gait control for given morphologies.

MINGUS Project-Team

4. Application Domains

4.1. Application domains

The MINGUS project aims at applying the new numerical methods on realistic problems arising for instance in physics of nanotechnology and physics of plasmas. Therefore, in addition to efforts devoted to the design and the analysis of numerical methods, the inherent large size of the problems at hand requires advanced mathematical and computational methods which are hard to implement. Another application is concerned with population dynamics for which the main goal is to understand how the spatial propagation phenomena affect the demography of a population (plankton, parasite fungi, ...). Our activity is mostly at an early stage in the process of transfer to industry. However, all the models we use are physically relevant and all have applications in many areas (ITER, Bose-Einstein condensate, wave turbulence, optical tomography, transport phenomena, population dynamics, ...). As a consequence, our research aims at reaching theoretical physicists or computational scientists in various fields who have strong links with industrial applications. In order to tackle as realistic physical problems as possible, a fundamental aspect will consist in working on the realization of numerical methods and algorithms which are able to make an efficient use of a large number of processors. Then, it is essential for the numerical methods developed in the MINGUS project to be thought through this prism. We will benefit from the strong expertise of P. Navaro in scientific computing and more precisely on the Selalib software library (see description below). Below, we detail our main applications: first, the modeling and numerical approximation of magnetized plasmas is our major application and will require important efforts in terms of software developments to scale-up our multiscale methods; second, the transport of charged particles in nanostructures has very interesting applications (like graphene material), for which our contributions will mainly focus on dedicated problems; lastly, applications on population dynamics will be dedicated to mathematical modeling and some numerical validations.

4.1.1. Plasmas problems

The Selalib (SEmi-LAgrangian LIBrary) software library⁰ is a modular library for kinetic and gyrokinetic simulations of plasmas in fusion energy devices. Selalib is a collection of fortran modules aimed at facilitating the development of kinetic simulations, particularly in the study of turbulence in fusion plasmas. Selalib offers basic capabilities and modules to help parallelization (both MPI and OpenMP), as well as pre-packaged simulations. Its main objective is to develop a documented library implementing several numerical methods for the numerical approximation of kinetic models. Another objective of the library is to provide physicists with easy-to-use gyrokinetic solvers. It has been originally developed by E. Sonnendrücker and his collaborators in the past CALVI Inria project, and has played an important role in the activities of the IPL FRATRES. P. Navaro is one of the main software engineer of this library and as such he played an important daily role in its development and its portability on supercomputers. Though Selalib has reached a certain maturity some additional works are needed to make available by the community. There are currently discussions for a possible evolution of Selalib, namely the writing of a new release which will be available for free download. Obviously, the team will be involved in this process. At the scientific level, Selalib is of great interest for us since it provides a powerful tool with which we can test, validate and compare our new methods and algorithms (users level). Besides numerical algorithms the library provides low-level utilities, input-output modules as well as parallelization strategies dedicated to kinetic problems. Moreover, a collection of simulations for typical test cases (of increasing difficulties) with various discretization schemes supplements the library. This library turns out to be the ideal complement of our activities and it will help us to scale-up our numerical methods to high-dimensional kinetic problems. During the last years, several experiments have been successfully performed in this direction (especially with PhD students) and it is important for us that this approach remains throughout. Then, we intend to integrate several of the numerical methods developed by the team within the Selalib library,

⁰SELALIB, <http://selalib.gforge.inria.fr>

with the strong help of P. Navaro (contributors level). This work has important advantages: (i) it will improve our research codes (in terms of efficiency but also of software maintenance point of view); (ii) it will help us to promote our research by making our methods available to the research community.

4.1.2. Quantum problems

Nowadays, a great challenge consists in the downscaling at the nanometer scale of electronic components in order to improve speed and efficiency of semiconductor materials. In this task, modeling and numerical simulations play an important role in the determination of the limit size of the nanotransistors. At the nanoscale, quantum effects have to be considered and the Schrödinger equation is prominent equation in this context. In the so-called semiclassical regime or when the transport is strongly confined, the solution endows space-time highly oscillations which are very difficult to capture numerically. An important application is the modeling of charged particles transport in graphene. Graphene is a sheet of carbone made of a single layer of molecule, organized in a bidimensional honeycomb crystal. The transport of charged particles in this structure is usually performed by Dirac equation (which is the relativistic counterpart of the Schrödinger equation). Due to the unusual properties of graphene -at room temperature, electrons moving in graphene behave as massless relativistic particles- physicists and compagnies are nowadays actively studying this material. Here, predicting how the material properties are affected by the uncertainties in the hexagonal lattice structure or in external potentials, is a major issue.

4.1.3. Population dynamics

The main goal is to characterize how spatial propagation phenomena (diffusion, transport, advection, . . .) affect the time evolution of the demography of a population. In collaboration with Y. Lagadeuc (ECOBIO, Rennes), this question has been studied for plankton. In this context, mathematical models have been proposed and it has been shown that the spatial dynamic (in this context, due to the marine current) which is fast compared to demographic scales, can strongly modify the demographic evolution of the plankton. In collaboration with Ecole d'Agronomie de Rennes, a mathematical study on the demography of a parasite fungi of plants has been performed. In this context, the demography is specific: the fungi can proliferate through sexual reproduction or through parthenogenesis. This two ways of reproduction give rise mathematically to quadratic and linear growth rates with respect to the population variable. The demography is then coupled with transport (transport of fungi spore by wind). Here, the goal is characterize the propagation of the fungi population by finding travelling waves solutions which are well adapted to describe the evolution of invasive fronts. Moreover, this approach enables to recover with a good agreement realistic examples (infection of ash or banana tree) for which experimental data are available. In these contexts, mathematical models are a powerful tool for biologists since measurements are very complicated to obtain and laboratory experiments hardly reproduce reality. The models derived are multiscale due to the nature of the underlying phenomena and the next step is to provide efficient numerical schemes.

Myriads Project-Team

4. Application Domains

4.1. Main application domains

The Myriads team investigates the design and implementation of system services. Thus its research activities address a broad range of application domains. We validate our research results with selected use cases in the following application domains:

- Smart city services,
- Smart grids,
- Energy and sustainable development,
- Home IoT applications,
- Bio-informatics applications,
- Data science applications,
- Computational science applications,
- Numerical simulations.

PACAP Project-Team

4. Application Domains

4.1. Domains

The PACAP team is working on the fundamental technologies for computer science: processor architecture, performance-oriented compilation and guaranteed response time for real-time. The research results may have impact on any application domain that requires high performance execution (telecommunication, multimedia, biology, health, engineering, environment...), but also on many embedded applications that exhibit other constraints such as power consumption, code size and guaranteed response time. Our research activity implies the development of software prototypes.

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.

4.4. Remote Sensing

Remote sensing is an image acquisition paradigm in which various pieces of information from the surface of the earth are obtained by airborne or satellite imaging. In the recent years the quality and variety of sensors available on the market for such remote acquisitions has drastically improved, leaving a large number of unsolved or difficult data processing tasks on the hand of data scientists. In particular, members of the PANAMA project-team have focused on processing hyperspectral images, that are images collected at a large number of wavelengths. The tasks that have been addressed are mainly spectral unmixing, that aims at identifying the various materials on the earth surface with their relative abundances, and denoising, that aims at removing structured noise from remote acquisitions that can be polluted by e.g. sensor movement or clouds. The spectra measured at each pixel of an hyperspectral image are heavily dependent on surface features (material, slop, material mixture type) which makes these data extremely rich and interesting.

RAINBOW Project-Team (section vide)

SERPICO Project-Team

4. Application Domains

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

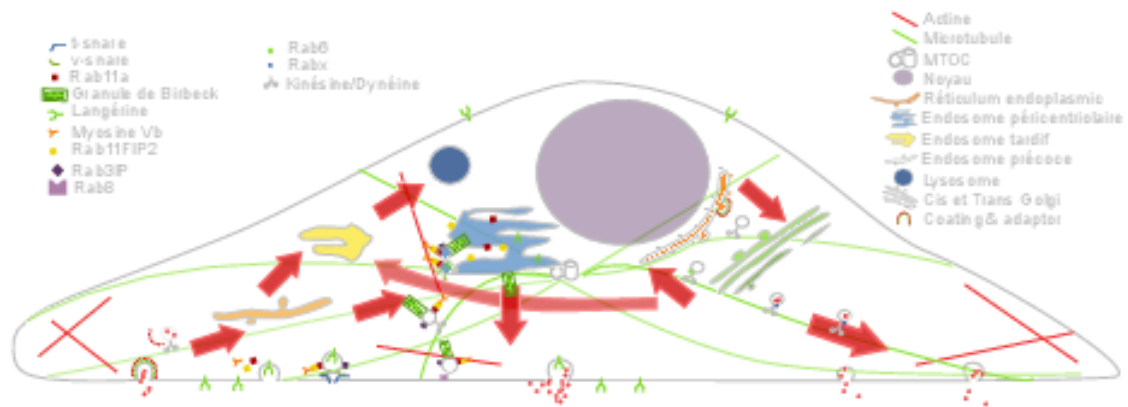


Figure 1. Cargo Langerin Trafficking controlled by Rab11A/Rab11FIP2/MyoVb platform.

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

In this context, the central aim of SERPICO is to understand how the different machineries of molecular components involved are interconnected and coordinated to generate such specialized structures, an issue that become more and more accessible, thanks to improvement in all domains related to live imaging. We need to address the following topics:

1. developing new bioimaging approaches to observe and statistically analyze such coordinated dynamics in live material;

2. correlating this statistically relevant spatiotemporal organization of protein networks with the biological architectures and at the ultrastructural level;
3. modeling intracellular transport of those reference biological complex systems and proposing new experimental plans in an iterative and virtuous circle;
4. managing and analyzing the workflow of image data obtained along different multidimensional microscopy modalities.

These studies are essential to unravel the complexity of the endomembrane system, how different machineries evolve together and thus coordinate (e.g. see Fig. 1). They help to decipher cell organization and function at different scales through an integrative workflow of methodological and technological developments. New approaches, such as optogenetics may even help controlling cell functions.

At long term, these studies will shed light on the cellular and molecular mechanisms underlying antigen presentation, viral infection or defense mechanisms, skin pigmentation, the pathogenesis of hereditary genetic disorders (lysosomal diseases, immune disorders) and on the mechanisms underlying cell differentiation and cell transformation. Our methodological goal is also to link dynamics information obtained through diffraction limited light microscopy, at a time regime compatible with live cell imaging and close to biochemical molecular interactions. The overview of ultrastructural organization will be achieved by complementary electron microscopy methods which have also undergone a revolutionary improvement over the last decade. Image visualization and quantitative analysis are of course essential issues in this context.

4.2. Imaging and analysis of cytoskeleton dynamics during cell migration

The ability to migrate in space is among the most fundamental functions of eukaryotic cells and thus is one of the best-studied phenomena in biology. During embryonic development, cell movements result in a massive reorganization of the embryo, from a simple spherical ball of cells into a multi-layered organism; many of the cells at or near the surface of the embryo move to a new, more interior location. Moreover, inadequate or inappropriate migration of immune cells is also critically important for the delivery of protective immune responses to tissues and for wound healing. Finally, cell migration may facilitate the dissemination of tumor cells from primary tumor in blood (extravasation) and eventually the colonization of other organs and the formation of secondary tumors.

It has been established that the cytoskeleton, composed of actin filaments, microtubules and intermediate filaments (elongated structures with a diameter of a few dozens of nanometers), is essential for several cell mechanisms, including cell migration, cell division and molecule trafficking:

- i/ the actin filaments promote cell protrusion, adhesion and retraction;
- ii/ the microtubules are the support of molecule traffic and cell polarization;
- iii/ the intermediate filaments are hypothesized to control microtubule organization.

Nevertheless, the mechanical and chemical states of migrating cells under various external conditions remain largely unknown. In the last decade, high-resolution microscopy methods led to the discovery of novel aspects of cell migration. Most approaches and models are limited to migration in 2D, justified by the flatness of the cell-motile mechanisms. However, the mechanical patterns that govern migration in 2D models are often not essential for efficient migration in 3D. Accordingly, recent very challenging 3D models of cells moving on flat surfaces have begun to emerge. The key challenge, however, is to understand how a 3D motile cell crawls through the 3D extracellular matrix. Another issue is of course to measure and understand how membrane protrusion and retraction keep the cell in homeostasis, which of course relate to membrane traffic.

The objective of SERPICO is to develop high-end signal processing and computer vision tools to unfold the dynamical coordination of microtubules, actin filaments and intermediate filaments in 3D, involved in cell migration, cell division and how molecular trafficking is coordinated with cytoskeleton changes in these fundamental cellular functions.

SIMSMART Project-Team

4. Application Domains

4.1. Domain 1 – Computational Physics

The development of large-scale computing facilities has enabled simulations of systems at the *atomistic scale* on a daily basis. The aim of these simulations is to bridge the time and space scales between the macroscopic properties of matter and the stochastic atomistic description. Typically, such simulations are based on the ordinary differential equations of classical mechanics supplemented with a random perturbation modeling temperature, or collisions between particles.

Let us give a few examples. In bio-chemistry, such simulations are key to predict the influence of a ligand on the behavior of a protein, with applications to drug design. The computer can thus be used as a *numerical microscope* in order to access data that would be very difficult and costly to obtain experimentally. In that case, a rare event (Objective 1) is given by a macroscopic system change such as a conformation change of the protein. In nuclear safety, such simulations are key to predict the transport of neutrons in nuclear plants, with application to assessing aging of concrete. In that case, a rare event is given by a high energy neutron impacting concrete containment structures.

A typical model used in molecular dynamics simulation of open systems at given temperature is a stochastic differential equation of Langevin type. The large time behavior of such systems is typically characterized by a hopping dynamics between 'metastable' configurations, usually defined by local minima of a potential energy. In order to bridge the time and space scales between the atomistic level and the macroscopic level, specific algorithms enforcing the realization of rare events have been developed. For instance, splitting particle methods (Objective 1) have become popular within the computational physics community only within the last few years, partially as a consequence of interactions between physicists and Inria mathematicians in ASPI (parent of SIMSMART) and MATHERIALS project-teams.

4.2. Domain 2 – Meteorology

The traditional trend in data assimilation in geophysical sciences (climate, meteorology) is to use as prior information some very complex deterministic models formulated in terms of fluid dynamics and reflecting as much as possible the underlying physical phenomenon (see *e.g.* <https://www.metoffice.gov.uk/research/modelling-systems/unified-model/>). Weather/climate forecasting can then be recast in terms of a Bayesian filtering problem (see Objective 2) using weather observations collected *in situ*.

The main issue is therefore to perform such Bayesian estimations with very expensive infinite dimensional prior models, and observations in large dimension. The use of some linear assumption in prior models (Kalman filtering) to filter non-linear hydrodynamical phenomena is the state-of-the-art approach, and a current field of research, but is plagued with intractable instabilities.

This context motivates two research trends: (i) the introduction of non-parametric, model-free prior dynamics constructed from a large amount of past, recorded real weather data; and (ii) the development of appropriate non-linear filtering approaches (Objective 2 and Objective 3).

SIMSMART will also test its new methods on multi-source data collected in North-Atlantic paying particular attention to coastal areas (*e.g.* within the inter-Labex SEACS).

SIROCCO Project-Team

4. Application Domains

4.1. Overview

The application domains addressed by the project are:

- Compression with advanced functionalities of various imaging modalities
- Networked multimedia applications taking into account needs in terms of user and network adaptation (e.g., interactive streaming, resilience to channel noise)
- Content editing, post-production, and computational photography.

4.2. Compression of emerging imaging modalities

Compression of visual content remains a widely-sought capability for a large number of applications. This is particularly true for mobile applications, as the need for wireless transmission capacity will significantly increase during the years to come. Hence, efficient compression tools are required to satisfy the trend towards mobile access to larger image resolutions and higher quality. A new impulse to research in video compression is also brought by the emergence of new imaging modalities, e.g. high dynamic range (HDR) images and videos (higher bit depth, extended colorimetric space), light fields and omni-directional imaging.

Different video data formats and technologies are envisaged for interactive and immersive 3D video applications using omni-directional videos, stereoscopic or multi-view videos. The "omni-directional video" set-up refers to 360-degree view from one single viewpoint or spherical video. Stereoscopic video is composed of two-view videos, the right and left images of the scene which, when combined, can recreate the depth aspect of the scene. A multi-view video refers to multiple video sequences captured by multiple video cameras and possibly by depth cameras. Associated with a view synthesis method, a multi-view video allows the generation of virtual views of the scene from any viewpoint. This property can be used in a large diversity of applications, including Three-Dimensional TV (3DTV), and Free Viewpoint Video (FVV). In parallel, the advent of a variety of heterogeneous delivery infrastructures has given momentum to extensive work on optimizing the end-to-end delivery QoS (Quality of Service). This encompasses compression capability but also capability for adapting the compressed streams to varying network conditions. The scalability of the video content compressed representation and its robustness to transmission impairments are thus important features for seamless adaptation to varying network conditions and to terminal capabilities.

4.3. Networked visual applications

Free-viewpoint Television (FTV) is a system for watching videos in which the user can choose its viewpoint freely and change it at anytime. To allow this navigation, many views are proposed and the user can navigate from one to the other. The goal of FTV is to propose an immersive sensation without the disadvantage of Three-dimensional television (3DTV). With FTV, a look-around effect is produced without any visual fatigue since the displayed images remain 2D. However, technical characteristics of FTV are large databases, huge numbers of users, and requests of subsets of the data, while the subset can be randomly chosen by the viewer. This requires the design of coding algorithms allowing such a random access to the pre-encoded and stored data which would preserve the compression performance of predictive coding. This research also finds applications in the context of Internet of Things in which the problem arises of optimally selecting both the number and the position of reference sensors and of compressing the captured data to be shared among a high number of users.

Broadband fixed and mobile access networks with different radio access technologies have enabled not only IPTV and Internet TV but also the emergence of mobile TV and mobile devices with internet capability. A major challenge for next internet TV or internet video remains to be able to deliver the increasing variety of media (including more and more bandwidth demanding media) with a sufficient end-to-end QoS (Quality of Service) and QoE (Quality of Experience).

4.4. Editing, post-production and computational photography

Editing and post-production are critical aspects in the audio-visual production process. Increased ways of “consuming” visual content also highlight the need for content repurposing as well as for higher interaction and editing capabilities. Content repurposing encompasses format conversion (retargeting), content summarization, and content editing. This processing requires powerful methods for extracting condensed video representations as well as powerful inpainting techniques. By providing advanced models, advanced video processing and image analysis tools, more visual effects, with more realism become possible. Our activities around light field imaging also find applications in computational photography which refers to the capability of creating photographic functionalities beyond what is possible with traditional cameras and processing tools.

STACK Project-Team

4. Application Domains

4.1. Overview

Supporting industrial actors and open-source communities in building an advanced software management stack is a key element to favor the advent of new kinds of information systems as well as web applications. Augmented reality, telemedicine and e-health services, smart-city, smart-factory, smart-transportation and remote security applications are under investigations. Although, STACK does not intend to address directly the development of such applications, understanding their requirements is critical to identify how the next generation of ICT infrastructures should evolve and what are the appropriate software abstractions for operators, developers and end-users. STACK team members have been exchanging since 2015 with a number of industrial groups (notably Orange Labs and Airbus), a few medical institutes (public and private ones) and several telecommunication operators in order to identify both opportunities and challenges in each of these domains, described hereafter.

4.2. Industrial Internet

The Industrial Internet domain gathers applications related to the convergence between the physical and the virtual world. This convergence has been made possible by the development of small, lightweight and cheap sensors as well as complex industrial physical machines that can be connected to the Internet. It is expected to improve most processes of daily life and decision processes in all societal domains, affecting all corresponding actors, be they individuals and user groups, large companies, SMEs or public institutions. The corresponding applications cover: the improvement of business processes of companies and the management of institutions (*e.g.*, accounting, marketing, cloud manufacturing, etc.); the development of large “smart” applications handling large amounts of geo-distributed data and a large set of resources (video analytics, augmented reality, etc.); the advent of future medical prevention and treatment techniques thanks to the intensive use of ICT systems, etc. We expect our contributions will favor the rise of efficient, correct and sustainable massively geo-distributed infrastructures that are mandatory to design and develop such applications.

4.3. Internet of Skills.

The Internet of Skills is an extension of the Industrial Internet to human activities. It can be seen as the ability to deliver physical experiences remotely (*i.e.*, via the Tactile Internet). Its main supporters advocate that it will revolutionize the way we teach, learn, and interact with pervasive resources. As most applications of the Internet of Skills are related to real time experiences, latency may be even more critical than for the Industrial Internet and raise the locality of computations and resources as a priority. In addition to identifying how Utility Computing infrastructures can cope with this requirement, it is important to determine how the quality of service of such applications should be defined and how latency and bandwidth constraints can be guaranteed at the infrastructure level.

4.4. e-Health

The e-Health domain constitutes an important societal application domain of the two previous areas. The STACK teams is investigating distribution, security and privacy issues in the fields of systems and personalized (aka. precision) medicine. The overall goal in these fields is the development of medication and treatment methods that are tailored towards small groups or even individual patients.

We are working, as part of the ongoing PrivGen CominLabs collaborative project on new means for the sharing of genetic data and applications in the Cloud. More generally, we are applying and developing corresponding techniques for the medical domains of genomics, immunobiology and transplantology in the international network SHLARC and the regional networks SysMics and Oncoshare: there, we investigate how to secure and preserve privacy if potentially sensitive personal data is moved and processed by distributed biomedical analyses.

We are also involved in the SyMeTRIC regional initiative where preliminary studies have been conducted in order to build a common System Medicine computing infrastructure to accelerate the discovery and validation of bio-markers in the fields of oncology, transplantation, and chronic cardiovascular diseases. The challenges were related to the need of being able to perform analyses on data that cannot be moved between distinct locations.

The STACK team will continue to contribute to the e-Health domain by harnessing advanced architectures, applications and infrastructures for the Fog/Edge.

4.5. Network Virtualization and Mobile Edge Services.

Telecom operators have been among the first to advocate the deployment of massively geo-distributed infrastructures, in particular through working groups such as Mobile Edge Computing at the European Telecommunication Standards Institute⁰. The initial reason is that geo-distributed infrastructures will enable Telecom operators to virtualize a large part of their resources and thus reduce capital and operational costs. As an example, we are investigating through the I/O Lab, the joint lab between Orange and Inria, how can a Centralized Radio Access Networks (*a.k.a.* C-RAN or Cloud-RAN) be supported for 5G networks. We highlight that our expertise is not on the network side but rather on where and how we can deploy, allocate and reconfigure software components, which are mandatory to operate a C-RAN infrastructure, in order to guarantee the quality of service expected by the end-users. Finally, working with actors from the network community is a valuable advantage for a distributed system research group such as STACK. Indeed, achievements made within one of the two communities serve the other.

⁰<http://www.etsi.org/news-events/news/1078-2016-04-etsi-mobile-edge-computing-publishes-foundation-specifications>.

SUMO Project-Team

4. Application Domains

4.1. Smart transportation systems

The smart-city trend aims at optimizing all functions of future cities with the help of digital technologies. We focus on the segment of urban trains, which will evolve from static and scheduled offers to reactive and eventually on-demand transportation offers. We address two challenges in this field. The first one concerns the optimal design of robust subway lines. The idea is to be able to evaluate, at design time, the performance of time tables and of different regulation policies. In particular, we focus on robustness issues: how can small perturbations and incidents be accommodated by the system, how fast will return to normality occur, when does the system become unstable? The second challenge concerns the design of new robust regulation strategies to optimize delays, recovery times, and energy consumption at the scale of a full subway line. These problems involve large-scale discrete-event systems, with temporal and stochastic features, and translate into robustness assessment, stability analysis and joint numerical/combinatorial optimization problems on the trajectories of these systems.

4.2. Management of telecommunication networks and of data centers

Telecommunication-network management is a rich provider of research topics for the team, and some members of SUMO have a long background of contacts and transfer with industry in this domain. Networks are typical examples of large distributed dynamic systems, and their management raises numerous problems ranging from diagnosis (or root-cause analysis), to optimization, reconfiguration, provisioning, planning, verification, etc. They also bring new challenges to the community, for example on the modeling side: building or learning a network model is a complex task, specifically because these models should reflect features like the layering, the multi-resolution view of components, the description of both functions, protocols and configuration, and they should also reflect dynamically-changing architectures. Besides modeling, management algorithms are also challenged by features like the size of systems, the need to work on abstractions, on partial models, on open systems, etc. The networking technology is now evolving toward software-defined networks, virtualized-network functions, multi-tenant systems, etc., which reinforces the need for more automation in the management of such systems.

Data centers are another example of large-scale modular dynamic and reconfigurable systems: they are composed of thousands of servers, on which virtual machines are activated, migrated, resized, etc. Their management covers issues like troubleshooting, reconfiguration, optimal control, in a setting where failures are frequent and mitigated by the performance of the management plane. We have a solid background in the coordination of the various autonomic managers that supervise the different functions/layers of such systems (hardware, middleware, web services, ...) Virtualization technologies now reach the domain of networking, and telecommunication operators/vendors evolve towards providers of distributed open clouds. This convergence of IT and networking strongly calls for new management paradigms, which is an opportunity for the team.

4.3. Collaborative workflows

A current trend is to involve end-users in collection and analysis of data. Examples of this trend are contributive science, crisis-management systems, and crowd sourcing applications. All these applications are data-centric and user-driven. They are often distributed and involve complex, and sometimes dynamic workflows. In many cases, there are strong interactions between data and control flows: indeed, decisions taken regarding the next tasks to be launched highly depend on collected data. For instance, in an epidemic-surveillance system, the aggregation of various reported disease cases may trigger alerts. Another example is crowd sourcing applications where user skills are used to complete tasks that are better performed by humans than computers.

In return, this requires addressing imprecise and sometimes unreliable answers. We address several issues related to complex workflows and data. We study declarative and dynamic models that can handle workflows, data, uncertainty, and competence management.

Once these models are mature enough, we plan to build prototypes to experiment them on real use cases from contributive science, health-management systems, and crowd sourcing applications. We also plan to define abstraction schemes allowing formal reasoning on these systems.

TAMIS Project-Team (section vide)

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 focuses on time in team TEA is central in any aspect of, cyber-physical, embedded system design in factory automation, automotive, music synthesis, signal processing, software radio, circuit and system on a chip design; many application domains which, should more collaborators join the team, would definitely be worth investigating.

Multi-scale, multi-aspect time modeling, analysis and software synthesis will greatly contribute to architecture modeling in these domains, with applications to optimized (distributed, parallel, multi-core) code generation for avionics (project Corail with Thales avionics, section 8) as well as modeling standards, real-time simulation and virtual integration in automotive (project with Toyota ITC, section 8).

Together with the importance of open-source software, one of these projects, the FUI Project P (section 8), demonstrated that a centralized model for system design could not just be a domain-specific programming language, such as discrete Simulink data-flows or a synchronous language. Synchronous languages implement a fixed model of time using logical clocks that are abstraction of time as sensed by software. They correspond to a fixed viewpoint in system design, and in a fixed hardware location in the system, which is not adequate to our purpose and must be extended.

In project P, we first tried to define a centralized model for importing discrete-continuous models onto a simplified implementation of SIMULINK: P models. Certified code generators would then be developed from that format. Because this does not encompass all aspects being translated to P, the P meta-model is now being extended to architecture description concepts (of the AADL) in order to become better suited for the purpose of system design. Another example is the development of System modeler on top of SCADE, which uses the more model-engineering flavored formalism SysML to try to unambiguously represent architectures around SCADE modules.

An abstract specification formalism, capable of representing time, timing relations, with which heterogeneous models can be abstracted, from which programs can be synthesized, naturally appears better suited for the purpose of virtual prototyping. RT-Builder, based on the data-flow language Signal and developed by TNI, was industrially proven and deployed for that purpose at Peugeot. It served to develop the virtual platform simulating all on-board electronics of PSA cars. This ‘hardware in the loop’ simulator was used to test equipments supplied by other manufacturers for virtual prototyping of cars. In the advent of the related automotive standard, RT-Builder then became AUTOSAR-Builder.

4.2. Factory Automation

In collaboration with Mitsubishi R&D, we explore another application domain where time and domain heterogeneity are prime concerns: factory automation. In factory automation alone, a system is conventionally built from generic computing modules: PLCs (Programmable Logic Controllers), connected to the environment with actuators and detectors, and linked to a distributed network. Each individual, physically distributed, PLC module must be timely programmed to perform individually coherent actions and fulfill the global physical, chemical, safety, power efficiency, performance and latency requirements of the whole production chain. Factory chains are subject to global and heterogeneous (physical, electronic, functional) requirements whose enforcement must be orchestrated for all individual components.

Model-based analysis in factory automation emerges from different scientific domains and focus on different CPS abstractions that interact in subtle ways: logic of PLC programs, real-time electro-mechanical processing, physical and chemical environments. This yields domain communication problems that render individual domain analysis useless. For instance, if one domain analysis (e.g. software) modifies a system model in a way that violates assumptions made by another domain (e.g. chemistry) then the detection of its violation may well be impossible to explain to either the software or chemistry experts. As a consequence, cross-domain analysis issues are discovered very late during system integration and lead to costly fixes. This is particularly prevalent in multi-tier industries, such as avionic, automotive, factories, where systems are prominently integrated from independently-developed parts.

WIDE Project-Team (section vide)