Activity Report 2018

Section Application Domains

COMETE Project-Team

4. Application Domains

4.1. Security and privacy

Participants: Catuscia Palamidessi, Konstantinos Chatzikokolakis, Ehab Elsalamouny, Ali Kassem, Anna Pazii, Marco Romanelli, Natasha Fernandes.

The aim of our research is the specification and verification of protocols used in mobile distributed systems, in particular security protocols. We are especially interested in protocols for *information hiding*.

Information hiding is a generic term which we use here to refer to the problem of preventing the disclosure of information which is supposed to be secret or confidential. The most prominent research areas which are concerned with this problem are those of *secure information flow* and of *privacy*.

Secure information flow refers to the problem of avoiding the so-called *propagation* of secret data due to their processing. It was initially considered as related to software, and the research focussed on type systems and other kind of static analysis to prevent dangerous operations, Nowadays the setting is more general, and a large part of the research effort is directed towards the investigation of probabilistic scenarios and treaths.

Privacy denotes the issue of preventing certain information to become publicly known. It may refer to the protection of *private data* (credit card number, personal info etc.), of the agent's identity (*anonymity*), of the link between information and user (*unlinkability*), of its activities (*unobservability*), and of its *mobility* (*untraceability*).

The common denominator of this class of problems is that an adversary can try to infer the private information (*secrets*) from the information that he can access (*observables*). The solution is then to obfuscate the link between secrets and observables as much as possible, and often the use randomization, i.e. the introduction of *noise*, can help to achieve this purpose. The system can then be seen as a *noisy channel*, in the information-theoretic sense, between the secrets and the observables.

We intend to explore the rich set of concepts and techniques in the fields of information theory and hypothesis testing to establish the foundations of quantitive information flow and of privacy, and to develop heuristics and methods to improve mechanisms for the protection of secret information. Our approach will be based on the specification of protocols in the probabilistic asynchronous π-calculus, and the application of model-checking to compute the matrices associated to the corresponding channels.

DATASHAPE Project-Team

4. Application Domains

4.1. Main application domains

Our work is mostly of a fundamental mathematical and algorithmic nature but finds a variety of applications in data analysis, e.g., in material science, biology, sensor networks, 3D shape analysis and processing, to name a few.

More specifically, DATASHAPE is working on the analysis of trajectories obtained from inertial sensors (PhD thesis of Bertrand Beaufils with Sysnav) and, more generally on the development of new TDA methods for Machine Learning and Artificial Intelligence for (multivariate) time-dependent data from various kinds of sensors in collaboration with Fujitsu.

<p style="text-align:center;color:red;">DEDUCTEAM Project-Team</p>

4. Application Domains

4.1. Interoperability

Our main impact applications, for instance to proofs of programs, or to air traffic control, are through our cooperation with other teams.

As a matter of fact, we view our work on interoperability and on the design of a formal proof encyclopedia as a service to the formal proof community.

GRACE Project-Team

4. Application Domains

4.1. Internet of Things

The *Internet of Things* (IoT) is the network and application space formed by the millions of small, connected devices that are increasingly present in our daily lives, and by the servers, clouds, and apps that they communicate with. This includes not only consumer devices such as smartphones, household devices, and wearable technology, but also an increasinly large proportion of our fundamental civic infrastructure (as is reflected by the increasing attention given to *Smart Cities*).

The IoT is therefore a massive, pervasive, and highly heterogeneous distributed computing system; a system that is mostly unprotected and insecure. Many of the devices are simply too small and underpowered to run the conventional cryptosystems that are standard for internet communications: even a minimalist TLS stack will often overwhelm the resources available on some small platforms. These limitations include small memory size, limited battery power, and low computational capacity. Not only are these devices harder to defend, but they are also much easier to attack: for example, these devices are generally extremely physically accessible (they must be, to fulfil their purpose), but this makes them extremely vulnerable to side-channel attacks.

Nevertheless, strong cryptography is essential to the future of IoT, precisely because these systems are so pervasive in our everyday lives, both individually (in our homes) and collectively (in our cities, industries, and urban infrastructure). We need strong cryptography to protect the personal and industrial data that these devices collect, process, and transmit; but we also need strong cryptography to ensure that devices and services can identify and authenticate themselves and each other with confidence. It is not enough to simply put secure systems in place; we must also develop reliable software update mechanisms, tailored to the needs and challenges of the IoT space.

While these technical challenges have been met, to some extent, for symmetric cryptosystems (which means that we have reasonable means of encrypting data and ensuring its integrity), they pose a massive problem for implementers of asymmetric cryptosystems (including key exchange, signatures, identification, and authentication). Efficient asymmetric cryptosystems have long been a research focus for GRACE, and our expertise in elliptic curve cryptosystems is of particular relevance for IoT, since these cryptosystems typically require the fewest memory and bandwidth resources.

Looking towards the future, the massive contemporary research effort in postquantum cryptosystems has so far mostly yielded systems even less-suited to IoT than conventional asymmetric systems are. Nevertheless, there is some hope that postquantum security can be brought to some IoT devices, and we are hopeful that GRACE's strength in isogeny-based cryptography will have an impact here.

4.2. Cloud storage

The team is concerned with several aspect of reliability and security of cloud storage, obtained mainly with tools from coding theory. On the privacy side, we build protocols for so-called Private Information Retrieval which enable a user to query a remote database for an entry, while not revealing his query. For instance, a user could query a service for stock quotes without revealing with company he is interested in. On the availability side, we study protocols for proofs of retrievability, which enable a user to get assurance that a huge file is still available on a remote server, with a low bandwith protocol which does not require to download the whole file. For instance, in a peer-to-peer distributed storage system, where nodes could be rewarded for storing data, they can be audited with proof of retrievability protocols to make sure they indeed hold the data.

We investigate these problems with algebraic coding theory, mainly codes with locality (locally decodable codes, locally recoverable codes, and so on).

4.3. Blockchains

The huge interest shown by companies for blockchains and cryptocurrencies have attracted the attention of mainstream industries for new, advanced uses of cryptographic, beyond confidentiality, integrity and authentication. In particular, zero-knowledge proofs, computation with encrypted data, etc, are now revealing their potential in the blockchain context. Team Grace is investigating two topics in these areas: secure multiparty computation and so-called "STARKS".

Secure multiparty computation enables several participants to compute a common function of data they each secretly own, without each participant revealing his data to the other participants. This area has seen great progress in recent years, and the cryptogaphic protocols are now mature enough for practical use. This topic is new to project-team Grace, and we will investigate it in the context of blockchains, through the lenses of use for private "smart contracts". A PhD student has been hired since October, funded by IRT System-X.

(ZK-)STARKS stands for "(Zero-Knowledge) Scalable Transparent ARguments of Knowledge", which can be zero knowledge or not. These techniques enable to have short probabilistic proof of correctness of program execution, which can be quicly checked by a verifier, without requiring the verifier to redo the computation again. This topic is close to the problem of computational integrity, and its theoretical foundations originate back to the 90's, which saw the formulation and proof of the celebrated PCP theorem. A protocol family equivalent of STARKS, "SNARKS", are well established, performant and promoted by the zerocash protocol for anomymous cryptocurrency (and also available in Ethereum), and STARKS are seen as a future replacement for SNARKS, overcoming the SNARKS problem of trusted setup. At the core of STARKS lie algebraic codes, mainly basic Reed-Solomon codes, and we will investigate replacement for the Reed-Solomon codes, to allow more performant (shorter) STARKS.

<h1 style="color:red; text-align:center">MEXICO Project-Team</h1>

4. Application Domains

4.1. Telecommunications

Participants: Stefan Haar, Serge Haddad.

Stefan Haar, Serge Haddad.

MExICo's research is motivated by problems of system management in several domains, such as:

- In the domain of service oriented computing, it is often necessary to insert some Web service into an existing orchestrated business process, e.g. to replace another component after failures. This requires to ensure, often actively, conformance to the interaction protocol. One therefore needs to synthesize adaptators for every component in order to steer its interaction with the surrounding processes.

- Still in the domain of telecommunications, the supervision of a network tends to move from out-of-band technology, with a fixed dedicated supervision infrastructure, to in-band supervision where the supervision process uses the supervised network itself. This new setting requires to revisit the existing supervision techniques using control and diagnosis tools.

Currently, we have no active cooperation on these subjects.

4.2. Biological Systems

Participants: Thomas Chatain, Matthias Függer, Stefan Haar, Serge Haddad, Stefan Schwoon.

We have begun in 2014 to examine concurrency issues in systems biology, and are currently enlarging the scope of our research's applications in this direction. To see the context, note that in recent years, a considerable shift of biologists' interest can be observed, from the mapping of static genotypes to gene expression, i.e. the processes in which genetic information is used in producing functional products. These processes are far from being uniquely determined by the gene itself, or even jointly with static properties of the environment; rather, regulation occurs throughout the expression processes, with specific mechanisms increasing or decreasing the production of various products, and thus modulating the outcome. These regulations are central in understanding cell fate (how does the cell differenciate ? Do mutations occur ? etc), and progress there hinges on our capacity to analyse, predict, monitor and control complex and variegated processes. We have applied Petri net unfolding techniques for the efficient computation of attractors in a regulatory network; that is, to identify strongly connected reachability components that correspond to stable evolutions, e.g. of a cell that differentiates into a specific functionality (or mutation). This constitutes the starting point of a broader research with Petri net unfolding techniques in regulation. In fact, the use of ordinary Petri nets for capturing regulatory network (RN) dynamics overcomes the limitations of traditional RN models : those impose e.g. Monotonicity properties in the influence that one factor had upon another, i.e. always increasing or always decreasing, and were thus unable to cover all actual behaviours. Rather, we follow the more refined model of boolean networks of automata, where the local states of the different factors jointly detemine which state transitions are possible. For these connectors, ordinary PNs constitute a first approximation, improving greatly over the literature but leaving room for improvement in terms of introducing more refined logical connectors. Future work thus involves transcending this class of PN models. Via unfoldings, one has access – provided efficient techniques are available – to all behaviours of the model, rather than over-or under-approximations as previously. This opens the way to efficiently searching in particular for determinants of the cell fate : which attractors are reachable from a given stage, and what are the factors that decide in favor of one or the other attractor, etc. Our current research focusses cellular reprogramming on the one hand, and **distributed algorithms in wild or synthetic biological systems** on the other.

The latter is a distributed algorithms' view on microbiological systems, both with the goal to model and analyze existing microbiological systems as distributed systems, and to design and implement distributed algorithms in synthesized microbiological systems. Envisioned major long-term goals are drug production and medical treatment via synthesized bacterial colonies. We are approaching our goal of a distributed algorithm's view of microbiological systems from several directions: (i) Timing plays a crucial role in microbiological systems. Similar to modern VLSI circuits, dominating loading effects and noise render classical delay models unfeasible. In previous work we showed limitations of current delay models and presented a class of new delay models, so called involution channels. In [26] we showed that involution channels are still in accordance with Newtonian physics, even in presence of noise. (ii) In [7] we analyzed metastability in circuits by a three-valued Kleene logic, presented a general technique to build circuits that can tolerate a certain degree of metastability at its inputs, and showed the presence of a computational hierarchy. Again, we expect metastability to play a crucial role in microbiological systems, as similar to modern VLSI circuits, loading effects are pronounced. (iii) We studied agreement problems in highly dynamic networks without stability guarantees [28], [27]. We expect such networks to occur in bacterial cultures where bacteria communicate by producing and sensing small signal molecules like AHL. Both works also have theoretically relevant implications: The work in [27] presents the first approximate agreement protocol in a multidimensional space with time complexity independent of the dimension, working also in presence of Byzantine faults. In [28] we proved a tight lower bound on convergence rates and time complexity of asymptotic and approximate agreement in dynamic and classical static fault models. (iv) We are currently working with Da-Jung Cho, Manish Kushwaha (INRA), and Thomas Nowak (LRI) on biological infection models for *E. coli* colonies and M13 phages.

4.3. Autonomous Vehicles

Participant: Serge Haddad.

The validation of safety properties is a crucial concern for the design of computer guided systems, in particular for automated transport systems Our approach consists in analyzing the interactions of a randomized environment (roads, cross-sections, etc.) with a vehicle controller.

<p style="text-align:center">**PARSIFAL Project-Team**</p>

4. Application Domains

4.1. Trustworthy implementations of theorem proving techniques

The production of real-world verified software has made it necessary to integrate results coming from different theorem provers in a single certification package. One approach to this integration task is by exchanging proof evidence and relying on a backend proof-checker.

Another approach to integration consists in re-implementing the theorem proving techniques as proof-search strategies, on an architecture that guarantees correctness.

Inference systems in general, and focused sequent calculi in particular, can serve as the basis of such an architecture, providing primitives for the exploration of the search space. These form a trusted *Application Programming Interface* that can be used to program and experiment various proof-search heuristics without worrying about correctness. No proof-checking is needed if one trusts the implementation of the API.

This approach has led to the development of the Psyche engine, and to its latest branch CDSAT.

Three major research directions are currently being explored, based on the above:

- The first one is about formulating automated reasoning techniques in terms of inference systems, so that they fit the approach described above. While this is rather standard for technique used in first-order Automated Theorem Provers (ATP), such as resolution, superposition, etc, this is much less standard in SMT-solving, the branch of automated reasoning that can natively handle reasoning in a combination of mathematical theories: the traditional techniques developed there usually organise the collaborations between different reasoning black boxes, whose opaque mechanisms less clearly connect to proof-theoretical inference systems. We are therefore investigating new foundations for reasoning in combinations of theories, expressed as fine-grained inference systems, and developed the *Conflict-Driven Satisfiability framework* for these foundations [13].

- The second one is about understanding how to deal with quantifiers in presence of one or more theories: On the one hand, traditional techniques for quantified problems, such as *unification* [29] or *quantifier elimination* are usually designed for either the empty theory or very specific theories. On the other hand, the industrial techniques for combining theories (Nelson-Oppen, Shostak, MCSAT [64], [69], [73], [53]) are designed for quantifier-free problems, and quantifiers there are dealt with incomplete *clause instantiation* methods or *trigger*-based techniques [41]. We are working on making the two approaches compatible.

- The above architecture's modular approach raises the question of how its different modules can safely cooperate (in terms of guaranteed correctness), while some of them are trusted and others are not. The issue is particularly acute if some of the techniques are run concurrently and exchange data at unpredictable times. For this we explore new solutions based on Milner's *LCF* [63]. In [47], we argued that our solutions in particular provide a way to fulfil the "Strategy Challenge for SMT-solving" set by De Moura and Passmore [74].

4.2. Principled computation for strong lambda-calculi

The application domain of the *cost models and abstract machines for functional programs* line of work—when *application* is intended in concrete terms—is the implementation of proof assistants.

Both functional languages and proof assistants rely on the λ-calculus has reference model. Functional languages are built on the *weak* λ-calculus (where evaluation does not enter function bodies) whose theory is simple and whose implementation has been widely explored in the last decades. Proof assistants instead require the full power of the *strong* λ-calculus, whose theory is more involved and whose implementation has mostly been neglected by the literature.

The study of reasonable cost models naturally leads to a refined theory of implementations, where different techniques and optimisations are classified depending on their complexity (with respect to the cost model). This direction is particularly relevant for the strong λ-calculus, for which most implementations are developed in a *ad-hoc* way.

The theoretical study in particular pointed out that all available proof assistants are implemented following unreasonable implementation schemas, where *unreasonable* here means with potentially exponential overhead with respect to the number of steps in the calculus.

Beniamino Accattoli collaborates with Bruno Barras—one of the implementors of *Coq*, the most used proof assistant—and Claudio Sacerdoti Coen—one of the implementors of *Matita*—in order to develop a fine theory of implementation for proof assistants.

If *applications* are intended also at a more theoretical level, the study of reasonable cost models is also applicable to the development of quantitative denotational semantics, to higher-order approaches to complexity theory, and to implicit computational complexity.

<div style="text-align:center; color:red;">**SPECFUN Project-Team**</div>

4. Application Domains

4.1. Computer Algebra in Mathematics

Our expertise in computer algebra and complexity-driven design of algebraic algorithms has applications in various domains, including:

- combinatorics, especially the study of combinatorial walks,
- theoretical computer science, like by the study of automatic sequences,
- number theory, by the analysis of the nature of so-called periods.

<div align="center">

TOCCATA Project-Team

</div>

4. Application Domains

4.1. Domain 1

The application domains we target involve safety-critical software, that is where a high-level guarantee of soundness of functional execution of the software is wanted. Currently our industrial collaborations mainly belong to the domain of transportation, including aeronautics, railroad, space flight, automotive.

Verification of C programs, Alt-Ergo at Airbus Transportation is the domain considered in the context of the ANR U3CAT project, led by CEA, in partnership with Airbus France, Dassault Aviation, Sagem Défense et Sécurité. It included proof of C programs via Frama-C/Jessie/Why, proof of floating-point programs [116], the use of the Alt-Ergo prover via CAVEAT tool (CEA) or Frama-C/WP. Within this context, we contributed to a qualification process of Alt-Ergo with Airbus industry: the technical documents (functional specifications and benchmark suite) have been accepted by Airbus, and these documents were submitted by Airbus to the certification authorities (DO-178B standard) in 2012. This action is continued in the new project Soprano.

Certified compilation, certified static analyzers Aeronautics is the main target of the Verasco project, led by Verimag, on the development of certified static analyzers, in partnership with Airbus. This is a follow-up of the transfer of the CompCert certified compiler (Inria team Gallium) to which we contributed to the support of floating-point computations [61].

Transfer to the community of Ada development The former FUI project Hi-Lite, led by Adacore company, introduced the use of Why3 and Alt-Ergo as back-end to SPARK2014, an environment for verification of Ada programs. This is applied to the domain of aerospace (Thales, EADS Astrium). At the very beginning of that project, Alt-Ergo was added in the Spark Pro toolset (predecessor of SPARK2014), developed by Altran-Praxis: Alt-Ergo can be used by customers as an alternate prover for automatically proving verification conditions. Its usage is described in the new edition of the Spark book [0] (Chapter "Advanced proof tools"). This action is continued in the new joint laboratory ProofInUse. A recent paper [69] provides an extensive list of applications of SPARK, a major one being the British air control management *iFacts*.

Transfer to the community of Atelier B In the current ANR project BWare, we investigate the use of Why3 and Alt-Ergo as an alternative back-end for checking proof obligations generated by *Atelier B*, whose main applications are railroad-related software [0], a collaboration with Mitsubishi Electric R&D Centre Europe (Rennes) (joint publication [121]) and ClearSy (Aix-en-Provence).

SMT-based Model-Checking: Cubicle S. Conchon (with A. Mebsout and F. Zaidi from VALS team at LRI) has a long-term collaboration with S. Krstic and A. Goel (Intel Strategic Cad Labs in Hillsboro, OR, USA) that aims in the development of the SMT-based model checker Cubicle (http://cubicle.lri.fr/) based on Alt-Ergo [118][5]. It is particularly targeted to the verification of concurrent programs and protocols.

---

[0]http://www.altran-praxis.com/book/
[0]http://www.methode-b.com/

COMMANDS Project-Team

4. Application Domains

4.1. Fuel saving by optimizing airplanes trajectories

We have a collaboration with the startup Safety Line on the optimization of trajectories for civil aircrafts. Key points include the reliable identification of the plane parameters (aerodynamic and thrust models) using data from the flight recorders, and the robust trajectory optimization of the climbing and cruise phases. We use both local (quasi-Newton interior-point algorithms) and global optimization tools (dynamic programming). The local method for the climb phase is in production and has been used for several hundreds of actual plane flights.

4.2. Hybrid vehicles

We have a collaboration with IFPEN on the energy management for hybrid vehicles. A significant direction is the analysis and classification of traffic data. More specifically, we focus on the traffic probability distribution in the (speed,torque) plane, with a time / space subdivision (road segments and timeframes).

4.3. Biological systems

2018 was the last year of IPL Algae in Silico in which we tackled the optimization of photobioreactors in turbid conditions. Our participation to IPL Cosy was strenghtened through the co-supervision of the PhD of E. Weill with team Lifeware (Inria and Pasteur), starting from september 2018, and focused on the oprimization of heterogenous populations of micro-organisms.

DEFI Project-Team

4. Application Domains

4.1. Radar and GPR applications

Conventional radar imaging techniques (ISAR, GPR, etc.) use backscattering data to image targets. The commonly used inversion algorithms are mainly based on the use of weak scattering approximations such as the Born or Kirchhoff approximation leading to very simple linear models, but at the expense of ignoring multiple scattering and polarization effects. The success of such an approach is evident in the wide use of synthetic aperture radar techniques.

However, the use of backscattering data makes 3-D imaging a very challenging problem (it is not even well understood theoretically) and as pointed out by Brett Borden in the context of airborne radar: "In recent years it has become quite apparent that the problems associated with radar target identification efforts will not vanish with the development of more sensitive radar receivers or increased signal-to-noise levels. In addition it has (slowly) been realized that greater amounts of data - or even additional "kinds" of radar data, such as added polarization or greatly extended bandwidth - will all suffer from the same basic limitations affiliated with incorrect model assumptions. Moreover, in the face of these problems it is important to ask how (and if) the complications associated with radar based automatic target recognition can be surmounted." This comment also applies to the more complex GPR problem.

Our research themes will incorporate the development, analysis and testing of several novel methods, such as sampling methods, level set methods or topological gradient methods, for ground penetrating radar application (imaging of urban infrastructures, landmines detection, underground waste deposits monitoring,) using multistatic data.

4.2. Biomedical imaging

Among emerging medical imaging techniques we are particularly interested in those using low to moderate frequency regimes. These include Microwave Tomography, Electrical Impedance Tomography and also the closely related Optical Tomography technique. They all have the advantage of being potentially safe and relatively cheap modalities and can also be used in complementarity with well established techniques such as X-ray computed tomography or Magnetic Resonance Imaging.

With these modalities tissues are differentiated and, consequentially can be imaged, based on differences in dielectric properties (some recent studies have proved that dielectric properties of biological tissues can be a strong indicator of the tissues functional and pathological conditions, for instance, tissue blood content, ischemia, infarction, hypoxia, malignancies, edema and others). The main challenge for these functionalities is to built a 3-D imaging algorithm capable of treating multi-static measurements to provide real-time images with highest (reasonably) expected resolutions and in a sufficiently robust way.

Another important biomedical application is brain imaging. We are for instance interested in the use of EEG and MEG techniques as complementary tools to MRI. They are applied for instance to localize epileptic centers or active zones (functional imaging). Here the problem is different and consists into performing passive imaging: the epileptic centers act as electrical sources and imaging is performed from measurements of induced currents. Incorporating the structure of the skull is primordial in improving the resolution of the imaging procedure. Doing this in a reasonably quick manner is still an active research area, and the use of asymptotic models would offer a promising solution to fix this issue.

4.3. Non destructive testing and parameter identification

One challenging problem in this vast area is the identification and imaging of defaults in anisotropic media. For instance this problem is of great importance in aeronautic constructions due to the growing use of composite materials. It also arises in applications linked with the evaluation of wood quality, like locating knots in timber in order to optimize timber-cutting in sawmills, or evaluating wood integrity before cutting trees. The anisotropy of the propagative media renders the analysis of diffracted waves more complex since one cannot only relies on the use of backscattered waves. Another difficulty comes from the fact that the micro-structure of the media is generally not well known a priori.

Our concern will be focused on the determination of qualitative information on the size of defaults and their physical properties rather than a complete imaging which for anisotropic media is in general impossible. For instance, in the case of homogeneous background, one can link the size of the inclusion and the index of refraction to the first eigenvalue of so-called interior transmission problem. These eigenvalues can be determined form the measured data and a rough localization of the default. Our goal is to extend this kind of idea to the cases where both the propagative media and the inclusion are anisotropic. The generalization to the case of cracks or screens has also to be investigated.

In the context of nuclear waste management many studies are conducted on the possibility of storing waste in a deep geological clay layer. To assess the reliability of such a storage without leakage it is necessary to have a precise knowledge of the porous media parameters (porosity, tortuosity, permeability, etc.). The large range of space and time scales involved in this process requires a high degree of precision as well as tight bounds on the uncertainties. Many physical experiments are conducted in situ which are designed for providing data for parameters identification. For example, the determination of the damaged zone (caused by excavation) around the repository area is of paramount importance since microcracks yield drastic changes in the permeability. Level set methods are a tool of choice for characterizing this damaged zone.

4.4. Diffusion MRI

In biological tissues, water is abundant and magnetic resonance imaging (MRI) exploits the magnetic property of the nucleus of the water proton. The imaging contrast (the variations in the grayscale in an image) in standard MRI can be from either proton density, T1 (spin-lattice) relaxation, or T2 (spin-spin) relaxation and the contrast in the image gives some information on the physiological properties of the biological tissue at different physical locations of the sample. The resolution of MRI is on the order of millimeters: the greyscale value shown in the imaging pixel represents the volume-averaged value taken over all the physical locations contained that pixel.

In diffusion MRI, the image contrast comes from a measure of the average distance the water molecules have moved (diffused) during a certain amount of time. The Pulsed Gradient Spin Echo (PGSE) sequence is a commonly used sequence of applied magnetic fields to encode the diffusion of water protons. The term 'pulsed' means that the magnetic fields are short in duration, an the term gradient means that the magnetic fields vary linearly in space along a particular direction. First, the water protons in tissue are labelled with nuclear spin at a precession frequency that varies as a function of the physical positions of the water molecules via the application of a pulsed (short in duration, lasting on the order of ten milliseconds) magnetic field. Because the precessing frequencies of the water molecules vary, the signal, which measures the aggregate phase of the water molecules, will be reduced due to phase cancellations. Some time (usually tens of milliseconds) after the first pulsed magnetic field, another pulsed magnetic field is applied to reverse the spins of the water molecules. The time between the applications of two pulsed magnetic fields is called the 'diffusion time'. If the water molecules have not moved during the diffusion time, the phase dispersion will be reversed, hence the signal loss will also be reversed, the signal is called refocused. However, if the molecules have moved during the diffusion time, the refocusing will be incomplete and the signal detected by the MRI scanner if weaker than if the water molecules have not moved. This lack of complete refocusing is called the signal attenuation and is the basis of the image contrast in DMRI. the pixels showning more signal attenuation is associated with further water displacement during the diffusion time, which may be linked to physiological factors, such as higher cell membrane permeability, larger cell sizes, higher extra-cellular volume fraction.

We model the nuclear magnetization of water protons in a sample due to diffusion-encoding magnetic fields by a multiple compartment Bloch-Torrey partial differential equation, which is a diffusive-type time-dependent PDE. The DMRI signal is the integral of the solution of the Bloch-Torrey PDE. In a homogeneous medium, the intrinsic diffusion coeffcient D will appear as the slope of the semi-log plot of the signal (in appropriate units). However, because during typical scanning times, 50-100ms, water molecules have had time to travel a diffusion distance which is long compared to the average size of the cells, the slope of the semi-log plot of the signal is in fact a measure of an 'effective' diffusion coefficient. In DMRI applications, this measured quantity is called the 'apparent diffusion coefficient' (ADC) and provides the most commonly used form the image contrast for DMRI. This ADC is closely related to the effective diffusion coefficient obtainable from mathematical homogenization theory.

4.5. Fluid flow applications

Specific actions are devoted to the problem of atmospheric reentry simulations. We focus on several aspects : i) on the development of innovative algorithms improving the prediction of hypersonic flows and including system uncertainties, ii) on the application of these methods to the atmospheric reentry of space vehicles for the control and the optimization of the trajectory, iii) on the debris reentry, which is of fundamental importance for NASA, CNES and ESA. Several works are already initiated with funding from CNES, Thales, and ASL. An ongoing activity concerns the design of the Thermal Protection System (TPS) that shields the spacecraft from aerothermal heating, generated by friction at the surface of the vehicle. The TPS is usually composed of different classes of materials, depending on the mission and the planned trajectory. One major issue is to model accurately the material response to ensure a safe design. High-fidelity material modeling for ablative materials has been developed by NASA, but a lot of work is still needed concerning the assessment of physical and modeling uncertainties during the design process. Our objective is to set up a predictive numerical tool to reliably estimate the response of ablative materials for different aerothermal conditions.

An important effort is dedicated to the simulation of fluids featuring complex thermodynamic behavior, in the context of two distinct projects: the VIPER project, funded by Aquitaine Region, and a project with CWI (Scientific Computing Group). Dense gases (DGs) are defined as single-phase vapors operating at temperatures and pressures conditions close to the saturation curve. The interest in studying complex dynamics of compressible dense gas flows comes from the potential technological advantages of using these fluids in energy conversion cycles, such as in Organic Rankine Cycles (ORCs) which used dense gases as energy converters for biomass fuels and low-grade heat from geothermal or industrial waste heat sources. Since these fluids feature large uncertainties in their estimated thermodynamic properties (critical properties, acentric factor, etc.), a meaningful numerical prediction of the performance must necessarily take into account these uncertainties. Other sources of uncertainties include, but are not limited to, the inlet boundary conditions which are often unknown in dense gases applications. Moreover, a robust optimization must also include the more generic uncertainty introduced by the machining tolerance in the construction of the turbine blades.

<p style="text-align:center; color:red;">**DISCO Project-Team**</p>

4. Application Domains

4.1. Analysis and Control of life sciences systems

The team is involved in life sciences applications. The two main lines are the analysis of bioreactors models and the modeling of cell dynamics in Acute Myeloblastic Leukemias (AML) in collaboration with St Antoine Hospital in Paris. A recent new subject is the modelling of Dengue epidemics.

4.2. Energy Management

The team is interested in Energy management and considers optimization and control problems in energy networks.

GAMMA3 Project-Team (section vide)

<p style="text-align:center;">**RANDOPT Team**</p>

4. Application Domains

4.1. Application Domains

Applications of black-box algorithms occur in various domains. Industry but also researchers in other academic domains have therefore a great need to apply black-box algorithms on a daily basis. We see this as a great source of motivation to design better methods. Applications not only allow us to backup our methods and understand what are the relevant features to solve a real-world problem but also help identify novel difficulties or set priorities in terms of algorithm design.

Asides from the two applications to Machine Learning that we detail below, we however do not target a specific application domain and we are interested in possible black-box applications stemming from various origins. This is for us intrinsic to the nature of the methods we develop that are general purpose algorithms. Hence our strategy with respect to applications can be seen as opportunistic and our main selection criteria when approached by colleagues who want to develop a collaboration around an application is whether we judge the application interesting: that is the application brings new challenges and/or gives us the opportunity to work on topics we already intended to work on.

The three concrete applications related to industrial collaborations we are currently dealing with are:

- With EDF R&D through the design and placement of bi-facial photovoltaic panel for the postdoc of Asma Atamna funded by the PGMO project.

- With Thales for the thesis of Konstantinos Varelas (DGA-CIFRE thesis) related to the design of radars (shape optimization of the wave form). This thesis investigates the development of large-scale variants of CMA-ES.

- With Storengy, a subsidiary of Engie specialized in gas storage for the thesis of Cheikh Touré. Different multi-objective applications are considered in this context but the primary motivation of Storengy is to get at their disposal a better multi-objective variant of CMA-ES which is the main objective of the developments within the thesis.

Additionally, there are two specific types of applications stemming from Machine Learning we would like to focus on: problems with non-differentiable loss that can occur in reinforcement learning and hyperparameter tuning problems. For the first class of problems the motivation comes from the paper [36] where different reinforcement learning problems are addressed and the weights of neural networks are adjusted using evolution strategies. Those problems are large-scale (in [36] up to 10^6 weights are adjusted), and the large-scale variants of CMA-ES we want to investigate might be relevant in this case. For the second class of problems (hyperparameter tuning problems), standard approches to handle those problems are Bayesian optimization algorithms but despite the tremendous effort for developing Bayesian optimization techniques and having implementations of Bayesian optimization algorithms within libraries, pure random search is still often used for training neural networks. One reason is that pure random search is intrinsically parallel [18]. This suggests that methods like CMA-ES—that are also intrinsically parallel—can be also advantageously used for hyperparameter tuning: this was demonstrated to tune deep neural networks in [32]. One limitation though of the CMA-ES algorithm is that it cannot deal with categorical/integer and continuous variables at the same time. This motivates us to investigate the development of CMA-ES variants that are able to deal with mixed variables.

When dealing with single applications, the results observed are difficult to generalize: typically not many methods are tested on a single application as tests are often time consuming and performed in restrictive settings. Yet, if one circumvent the problem of confidentiality of data and of criticality for companies to publish their applications, real-world problems could become benchmarks as any other analytical function. This would allow to test wider ranges of methods on the problems and to find out whether analytical benchmarks properly

capture real-world problem difficulties. We will thus seek to incorporate real-world problems within the COCO platform. This is a recurrent demand by researchers in optimization. As far as confidentiality of data are concerned, our preliminary discussions with industrials allow us to be optimistic that we can convince industrials to propose real-world problems with anonymized (and uncritical) data that still capture the essence of the underlying real-world problem.

SELECT Project-Team

4. Application Domains

4.1. Introduction

A key goal of SELECT is to produce methodological contributions in statistics. For this reason, the SELECT team works with applications that serve as an important source of interesting practical problems and require innovative methodology to address them. Many of our applications involve contracts with industrial partners, e.g., in reliability, although we also have several academic collaborations, e.g., in genetics and image analysis.

4.2. Curve classification

The field of classification for complex data such as curves, functions, spectra and time series, is an important problem in current research. Standard data analysis questions are being looked into anew, in order to define novel strategies that take the functional nature of such data into account. Functional data analysis addresses a variety of applied problems, including longitudinal studies, analysis of fMRI data, and spectral calibration.

We are focused in particular on unsupervised classification. In addition to standard questions such as the choice of the number of clusters, the norm for measuring the distance between two observations, and vectors for representing clusters, we must also address a major computational problem: the functional nature of the data, which requires new approaches.

4.3. Computer experiments and reliability

For several years now, SELECT has collaborated with the EDF-DER *Maintenance des Risques Industriels* group. One important theme involves the resolution of inverse problems using simulation tools to analyze incertainty in highly complex physical systems.

The other major theme concerns reliability, through a research collaboration with Nexter involving a Cifre convention. This collaboration concerns a lifetime analysis of a vehicle fleet to assess ageing.

Moreover, a collaboration is ongoing with Dassault Aviation on the modal analysis of mechanical structures, which aims to identify the vibration behavior of structures under dynamic excitation. From the algorithmic point of view, modal analysis amounts to estimation in parametric models on the basis of measured excitations and structural response data. In literature and existing implementations, the model selection problem associated with this estimation is currently treated by a rather weighty and heuristic procedure. In the context of our own research, model selection via penalization methods are being tested on this model selection problem.

4.4. Analysis of genomic data

For many years now, SELECT collaborates with Marie-Laure Martin-Magniette (URGV) for the analysis of genomic data. An important theme of this collaboration is using statistically sound model-based clustering methods to discover groups of co-expressed genes from microarray and high-throughput sequencing data. In particular, identifying biological entities that share similar profiles across several treatment conditions, such as co-expressed genes, may help identify groups of genes that are involved in the same biological processes.

SELECT collaborates with Anavaj Sakuntabhai and Philippe Dussart (Pasteur Institute) on predicting dengue severity using only low-dimensional clinical data obtained at hospital arrival. Further collaborations are underway in dengue fever and encephalitis with researchers at the Pasteur Institute, including with Jean-David Pommier.

SELECT is involved in the ANR "jeunes chercheurs" MixStatSeq directed by Cathy Maugis (INSA Toulouse), which is concerned with statistical analysis and clustering of RNASeq genomics data.

4.5. Pharmacovigilance

A collaboration is ongoing with Pascale Tubert-Bitter, Ismael Ahmed and Mohamed Sedki (Pharmacoepidemiology and Infectious Diseases, PhEMI) for the analysis of pharmacovigilance data. In this framework, the goal is to detect, as soon as possible, potential associations between certain drugs and adverse effects, which appeared after the authorized marketing of these drugs. Instead of working on aggregate data (contingency table) like is usually the case, the approach developed aims to deal with individual's data, which perhaps gives more information.

4.6. Spectroscopic imaging analysis of ancient materials

Ancient materials, encountered in archaeology and paleontology are often complex, heterogeneous and poorly characterized before physico-chemical analysis. A popular technique to gather as much physico-chemical information as possible, is spectro-microscopy or spectral imaging, where a full spectra, made of more than a thousand samples, is measured for each pixel. The produced data is tensorial with two or three spatial dimensions and one or more spectral dimensions, and requires the combination of an "image" approach with a "curve analysis" approach. Since 2010 SELECT, collaborates with Serge Cohen (IPANEMA) on clustering problems, taking spatial constraints into account.

4. Application Domains

4.1. Computational Social Sciences

Participants: Philippe Caillou, Isabelle Guyon, Michèle Sebag, Paola Tubaro
Collaboration: Jean-Pierre Nadal (EHESS); Marco Cuturi, Bruno Crépon (ENSAE); Thierry Weil (Mines); Jean-Luc Bazet (RITM)

Computational Social Sciences (CSS) studies social and economic phenomena, ranging from technological innovation to politics, from media to social networks, from human resources to education, from inequalities to health. It combines perspectives from different scientific disciplines, building upon the tradition of computer simulation and modeling of complex social systems [99] on the one hand, and data science on the other hand, fueled by the capacity to collect and analyze massive amounts of digital data.

The emerging field of CSS raises formidable challenges along three dimensions. Firstly, the definition of the research questions, the formulation of hypotheses and the validation of the results require a tight pluridisciplinary interaction and dialogue between researchers from different backgrounds. Secondly, the development of CSS is a touchstone for ethical AI. On the one hand, CSS gains ground in major, data-rich private companies; on the other hand, public researchers around the world are engaging in an effort to use it for the benefit of society as a whole [119]. The key technical difficulties related to data and model biases, and to self-fulfilling prophecies have been discussed in section 3.1 . Thirdly, CSS does not only regard scientists: it is essential that the civil society participate in the science of society [146].

TAO was involved in CSS for the last five years, and its activities have been strengthened thanks to P. Tubaro's and I. Guyon's expertises respectively in sociology and economics, and in causal modeling. Details are given in Section 7.3 .

4.2. Energy Management

Participants: Isabelle Guyon, Marc Schoenauer, Michèle Sebag
PhD: Victor Berger, Benjamin Donnot, Balthazar Donon, Herilalaina Rakotoarison
Collaboration: Antoine Marot, Patrick Panciatici (RTE), Vincent Renault (Artelys), Olivier Teytaud (Facebook)

Energy Management has been an application domain of choice for TAO since the end 2000s, with main partners SME Artelys (METIS Ilab Inria; ADEME project POST; ADEME project NEXT) and RTE (See.4C European challenge; two CIFRE PhDs). The goals concern i) optimal planning over several spatio-temporal scales, from investments on continental Europe/North Africa grid at the decade scale (POST), to daily planning of local or regional power networks (NEXT); ii) monitoring and control of the French grid enforcing the prevention of power breaks (RTE); iii) improvement of house-made numerical methods using data-intense learning (as described in Section 3.2) in all aspects of IFPEN activities, from geological problems in oil prospection (IF**P**EN) to the optimal placement of eolians in eolian fields (IF**PEN**).
Optimal planning over long periods of time amounts to optimal sequential decision under high uncertainties, ranging from stochastic uncertainties (weather, market prices, demand prediction) handled based on massive data, to non-stochastic uncertainties (e.g., political decisions about the nuclear policy) handled through defining and selecting a tractable number of scenarios. Note that non-anticipativity constraints forbid the use of dynamic programming-related methods; this led to propose the *Direct Value Search* method [79] at the end of the POST project. A further recent work in the same direction [21] proposes and theoretically studies the *Direct Model Predictive Control* approach, a hybrid model which merges the properties of two different dynamic optimization methods, Model Predictive Control and Stochastic Dual Dynamic Programming, has robust convergence properties, and experimentally competes with both methods alone.

The daily maintainance of power grids requires the building of approximate predictive models on the top of any given network topology. Deep Networks are natural candidates for such modelling, considering the size of the French grid ($\sim$ 10000 nodes), but the representation of the topology is a challenge when, e.g. the RTE goal is to quickly ensure the "n-1" security constraint (the network should remain safe even if any of the 10000 nodes fails). Existing simulators are too slow to be used in real time, and the size of actual grids makes it intractable to train surrogate models for all possible (n-1) topologies (see Section 7.4 for more details).

Even when efficient simulators do exist, they need to be calibrated (adjusting their hyper-parameters with real data), and complemented by uncertainty propagation models. Such adaptations and extensions are at the core of the NEXT project; hyper-parameter tuning is also a challenge regarding the development plans of the local grids, that heavily rely on graph optimization algorithms.

Furthermore, predictive models of local grids are based on the estimated consumption of end-customers: Linky meters provide coarse grain information only due to privacy issues, and very few samples of fine-grained consumption are available (from volunteer customers). A first task is to transfer knowledge from small data to the whole domain of application. A second task is to directly predict the peak of consumption based on the user cluster profiles and their representativity (see Section 7.4.2).

Another research direction formulates security maintenance as a reinforcement problem, taking inspiration from the recent successes of Deep Reinforcement Learning. This direction is being investigated in Balthazar Donon's RTE CIFRE PhD with RTE (started Oct. 2018).

4.3. Data-driven Numerical Modeling

Participants: Guillaume Charpiat, Cécile Germain, Isabelle Guyon, Flora Jay, Marc Schoenauer, Michèle Sebag
PhD and Post-doc: Victor Estrade, Loris Felardo, Adrian Pol, Théophile Sanchez
Collaboration: D. Rousseau (LAL), M. Pierini (CERN)

As said (section 3.2), in domains where both first principle-based models and equations, and empirical or simulated data are available, their combined usage can support more accurate modelling and prediction, and when appropriate, optimization, control and design. This section describes such applications, with the goal of improving the time-to-design chain through fast interactions between the simulation, optimization, control and design stages. The expected advances regard: i) the quality of the models or simulators (through data assimilation, e.g. coupling first principles and data, or repairing/extending closed-form models); ii) the exploitation of data derived from different distributions and/or related phenomenons; and, most interestingly, iii) the task of optimal design and the assessment of the resulting designs.

The proposed approaches are based on generative and adversarial modelling [116], [102], extending both the generator and the discriminator modules to take advantage of the domain knowledge.

A first challenge regards the design of the model space, and the architecture used to enforce the known domain properties (symmetries, invariance operators, temporal structures). When appropriate, data from different distributions (e.g. simulated vs real-world data) will be reconciled, for instance taking inspiration from real-valued non-volume preserving transformations [84] in order to preserve the natural interpretation.

Another challenge regards the validation of the models and solutions of the optimal design problems. The more flexible the models, the more intensive the validation must be, as reminded by Leon Bottou. Along this way, generative models will be used to support the design of "what if" scenarios, to enhance anomaly detection and monitoring via refined likelihood criteria.

<p style="text-align:center">**TROPICAL Project-Team**</p>

4. Application Domains

4.1. Discrete event systems (manufacturing systems, networks)

One important class of applications of max-plus algebra comes from discrete event dynamical systems [62]. In particular, modelling timed systems subject to synchronization and concurrency phenomena leads to studying dynamical systems that are non-smooth, but which have remarkable structural properties (nonexpansiveness in certain metrics , monotonicity) or combinatorial properties. Algebraic methods allow one to obtain analytical expressions for performance measures (throughput, waiting time, etc). A recent application, to emergency call centers, can be found in [55].

4.2. Optimal control and games

Optimal control and game theory have numerous well established applications fields: mathematical economy and finance, stock optimization, optimization of networks, decision making, etc. In most of these applications, one needs either to derive analytical or qualitative properties of solutions, or design exact or approximation algorithms adapted to large scale problems.

4.3. Operations Research

We develop, or have developed, several aspects of operations research, including the application of stochastic control to optimal pricing, optimal measurement in networks [122]. Applications of tropical methods arise in particular from discrete optimization [68], [69], scheduling problems with and-or constraints [114], or product mix auctions [130].

4.4. Computing program and dynamical systems invariants

A number of programs and systems verification questions, in which safety considerations are involved, reduce to computing invariant subsets of dynamical systems. This approach appears in various guises in computer science, for instance in static analysis of program by abstract interpretation, along the lines of P. and R. Cousot [75], but also in control (eg, computing safety regions by solving Isaacs PDEs). These invariant sets are often sought in some tractable effective class: ellipsoids, polyhedra, parametric classes of polyhedra with a controlled complexity (the so called "templates" introduced by Sankaranarayanan, Sipma and Manna [123]), shadows of sets represented by linear matrix inequalities, disjunctive constraints represented by tropical polyhedra [57], etc. The computation of invariants boils down to solving large scale fixed point problems. The latter are of the same nature as the ones encountered in the theory of zero-sum games, and so, the techniques developed in the previous research directions (especially methods of monotonicity, nonexpansiveness, discretization of PDEs, etc) apply to the present setting, see e.g. [90], [95] for the application of policy iteration type algorithms, or for the application for fixed point problems over the space of quadratic forms [7]. The problem of computation of invariants is indeed a key issue needing the methods of several fields: convex and nonconvex programming, semidefinite programming and symbolic computation (to handle semialgebraic invariants), nonlinear fixed point theory, approximation theory, tropical methods (to handle disjunctions), and formal proof (to certify numerical invariants or inequalities).

<p style="text-align:center">**GALEN-POST Team**</p>

4. Application Domains

4.1. Representation Learning for Network Biology

Participants: Fragkiskos Malliaros, Abdulkadir Çelikkanat (Collaboration: Duong Nguyen, UC San Diego)
Networks (or graphs) are ubiquitous in the domain of biology, as many biological systems can naturally be mapped to graph structures. Characteristic examples include protein-protein interaction and gene regulatory networks. To this extend, machine learning on graphs is an important task with many practical applications in network biology. For example, in the case on protein-protein interaction networks, predicting the function of a protein is a key task that assigns biochemical roles to proteins. The main challenge here is to find appropriate representations of the graph structure, in order to be easily exploited by machine learning models. The traditional approach to the problem was relying on the extraction of "hand-crafted" discriminating features that encode information about the graph, based on user-defined heuristics. Nevertheless, this approach has demonstrated severe limitations, as the learning process heavily depends on the manually extracted features. To this end, feature (or representation) learning techniques can be used to automatically learn to encode the graph structure into low-dimensional feature vectors – which can later be used in learning tasks. Our goal here is to develop a systematic framework for large-scale representation learning on biological graphs. Our approach takes advantage of the clustering structure of these networks, to further enhance the ability of the learned features to capture intrinsic structural properties.

4.2. Breast tomosynthesis

Participants: Emilie Chouzenoux, Jean-Christophe Pesquet, Maissa Sghaier (collaboration G. Palma, GE Healthcare)
Breast cancer is the most frequently diagnosed cancer for women. Mammography is the most used imagery tool for detecting and diagnosing this type of cancer. Since it consists of a 2D projection method, this technique is sensitive to geometrical limitations such as the superimposition of tissues which may reduce the visibility of lesions or make even appear false structures which are interpreted by radiologists as suspicious signs. Digital breast tomosynthesis allows these limitations to be circumvented. This technique is grounded on the acquisition of a set of projections with a limited angle view. Then, a 3D estimation of the sensed object is performed from this set of projections, so reducing the overlap of structures and improving the visibility and detectability of lesions possibly present in the breast. The objective of our work is to develop a high quality reconstruction methodology where the full pipeline of data processing will be modeled.

4.3. Inference of gene regulatory networks

Participants: Jean-Christophe Pesquet (collaboration A. Pirayre and L. Duval, IFPEN)
The discovery of novel gene regulatory processes improves the understanding of cell phenotypic responses to external stimuli for many biological applications, such as medicine, environment or biotechnologies. To this purpose, transcriptomic data are generated and analyzed from DNA microarrays or more recently RNAseq experiments. They consist in genetic expression level sequences obtained for all genes of a studied organism placed in dierent living conditions. From these data, gene regulation mechanisms can be recovered by revealing topological links encoded in graphs. In regulatory graphs, nodes correspond to genes. A link between two nodes is identified if a regulation relationship exists between the two corresponding genes. In our work, we propose to address this network inference problem with recently developed techniques pertaining to graph optimization. Given all the pairwise gene regulation information available, we propose to determine the presence of edges in the considered GRN by adopting an energy optimization formulation integrating additional constraints. Either biological (information about gene interactions) or structural (information about node connectivity) a priori are considered to restrict the space of possible solutions. Different priors lead to different properties of the global cost function, for which various optimization strategies, either discrete and continuous, can be applied.

4.4. Imaging biomarkers and characterization for chronic lung diseases

Participants: Guillaume Chassagnon, Maria Vakalopoulou (in collaboration with Evangelia Zacharaki and Nikos Paragios: University of Patras; Therapanacea)

Diagnosis and staging of chronic lung diseases is a major challenge for both patient care and approval of new treatments. Among imaging techniques, computed tomography (CT) is the gold standard for in vivo morphological assessment of lung parenchyma currently offering the highest spatial resolution in chronic lung diseases. Although CT is widely used its optimal use in clinical practice and as an endpoint in clinical trials remains controversial. Our goal is to develop quantitative imaging biomarkers that allow (i)severity assessment (based on the correlation to functional and clinical data) and (ii) monitoring the disease progression. In the current analysis we focus on scleroderma and cystic fibrosis as models for restrictive and obstructive lung disease, respectively. Two different approaches are investigated: disease assessment by deep convolutional neural networks and assessment of the regional lung elasticity through deformable registration. This work is in collaboration with the Department of Radiology, Cochin Hospital, Paris.

4.5. Imaging radiomics and genes to assess immunotherapy

Participants: Roger Sun, Théo Estienne, Enzo Batistella, Maria Vakalopoulou (in collaboration with Éric Deutsch and Nikos Paragios: Institut de Cancérologie Gustave Roussy, Therapanacea)

Because responses of patients with cancer to immunotherapy can vary in success, innovative predictors of response to treatment are urgently needed to improve treatment outcomes. We aimed to develop and independently validate a radiomics-based biomarker of tumour-infiltrating CD8 cells in patients included in phase 1 trials of anti-programmed cell death protein (PD)-1 or anti-programmed cell death ligand 1 (PD-L1) monotherapy. We also aimed to evaluate the association between the biomarker, and tumour immune phenotype and clinical outcomes of these patients. This work is in collaboration with the Institut de Cancérologie Gustave Roussy Paris.

4.6. Restoration of old video archives

Participants: Emilie Chouzenoux, Jean-Christophe Pesquet (collaboration F. Abboud, WITBE, J.-H. Chenot and L. Laborelli, INA)

The last century has witnessed an explosion in the amount of video data stored with holders such as the National Audiovisual Institute whose mission is to preserve and promote the content of French broadcast programs. the cultural impact of these records, their value is increased due to commercial reexploitation through recent visual media. However, the perceived quality of the old data fails to satisfy the current public demand. The purpose of our work is to propose new methods for restoring video sequences supplied from television archive documents, using modern optimization techniques with proven convergence properties [21], [50], [3].

4.7. Development of a heart ventricle vessel generation model for perfusion analysis

Participant: Hugues Talbot (collaboration with L. Najman ESIEE Paris, I. Vignon-Clementel, REO Team, Inria, Leo Grady, Heartflow Inc.)

Cardio-vascular diseases are the leading cause of mortality in the world. Understanding these diseases is still a current, challenging and essential research project. The leading cause of heart malfunction are stenoses causing ischemia in the coronary vessels. Current CT and MRI technology allow to assess coronary diseases but are typically invasive, requiring catheterization and relatively toxic contrast agents injection. With Heartflow, a US based company, we have in the past worked to use image-based exams only, limiting the use of contrast agents and in many cases eliminating catheterisation. Heartflow is current the market leader in non-invasive coronary exams.

Unfortunately, current imaging technology is unable to assess the full length of coronary vessels. CT is limited to a resolution of about 1mm, whereas coronary vessels can be much smaller. Blood perfusion throughout the heart muscle can provide insight regarding coronary health in areas that CT or MRI cannot assess. Perfusion imaging with PET or a Gamma camera, the current gold standard, is an invasive technology requiring the use of radioactive tracers.

With heartflow and with the REO team of Inria, we have investigated patient-specific vessel generation models together with porous model simulations in order to propose a forward model of perfusion imaging, based on the known patient data, computer flow dynamic simulations as well as experimental data consistent with known vessel and heart muscle physiology. The objective of this work is to solve the inverse problem of locating and assessing coronary diseases even though the affected vessels are too small to be imaged directly.

LIFEWARE Project-Team

4. Application Domains

4.1. Preamble

Our collaborative work on biological applications is expected to serve as a basis for groundbreaking advances in cell functioning understanding, cell monitoring and control, and novel therapy design and optimization. Our collaborations with biologists are focused on **concrete biological questions**, and on the building of predictive models of biological systems to answer them. Furthermore, one important application of our research is the development of a **modeling software** for computational systems biology.

4.2. Modeling software for systems biology and synthetic biology

Since 2002, we develop an open-source software environment for modeling and analyzing biochemical reaction systems. This software, called the Biochemical Abstract Machine (BIOCHAM), is compatible with SBML for importing and exporting models from repositories such as BioModels. It can perform a variety of static analyses, specify behaviors in Boolean or quantitative temporal logics, search parameter values satisfying temporal constraints, and make various simulations. While the primary reason of this development effort is to be able to **implement our ideas and experiment them quickly on a large scale**, BIOCHAM is used by other groups either for building models, for comparing techniques, or for teaching (see statistics in software section). BIOCHAM-WEB is a web application which makes it possible to use BIOCHAM without any installation. We plan to continue developing BIOCHAM for these different purposes and improve the software quality.

4.3. Coupled models of the cell cycle and the circadian clock

Recent advances in cancer chronotherapy techniques support the evidence that there exist important links between the cell cycle and the circadian clock genes. One purpose for modeling these links is to better understand how to efficiently target malignant cells depending on the phase of the day and patient characterictics. These questions are at the heart of our collaboration with Franck Delaunay (CNRS Nice) and Francis Lévi (Univ. Warwick, GB, formerly INSERM Hopital Paul Brousse, Villejuif) and of our participation in the ANR HYCLOCK project and in the submitted EU H2020 C2SyM proposal, following the former EU EraNet Sysbio C5SYS and FP6 TEMPO projects. In the past, we developed a coupled model of the Cell Cycle, Circadian Clock, DNA Repair System, Irinotecan Metabolism and Exposure Control under Temporal Logic Constraints [0]. We now focus on the bidirectional coupling between the cell cycle and the circadian clock and expect to gain fundamental insights on this complex coupling from computational modeling and single-cell experiments.

4.4. Biosensor design and implementation in non-living protocells

In collaboration with Franck Molina (CNRS, Sys2Diag, Montpellier) and Jie-Hong Jiang (NTU, Taiwan) we ambition to apply our techniques to the design and implementation of biosensors in non-living vesicles for medical applications. Our approach is based on purely protein computation and on our ability to compile controllers and programs in biochemical reactions. The realization will be prototyped using a microfluidic device at CNRS Sys2Diag which will allow us to precisely control the size of the vesicles and the concentrations of the injected proteins. It is worth noting that the choice of non-living chassis, in contrast to living cells in synthetic biology, is particularly appealing for security considerations and compliance to forthcoming EU regulation.

---

[0]Elisabetta De Maria, François Fages, Aurélien Rizk, Sylvain Soliman. Design, Optimization, and Predictions of a Coupled Model of the Cell Cycle, Circadian Clock, DNA Repair System, Irinotecan Metabolism and Exposure Control under Temporal Logic Constraints. Theoretical Computer Science, 412(21):2108 2127, 2011.

4.5. Functional characterization of the resistance of bacterial populations to antimicrobial treatments

Antibiotic resistance is becoming a problem of central importance at a global level. Two mechanisms are at the origin of non-susceptibility to antimicrobial treatments. The first one comes from adaptation of bacterial cells to antibacterial treatments, notably through the modification of efflux pumps or the expression of enzymes that degrade the antibiotics. Cells are individually resistant. The second one, typically found in resistances to β-lactams, a broad class of antibiotics, originates from the release in the environment of the antibiotic degrading enzymes by the dead cells. This leads to population effects by which cells become collectively resilient.

The functional characterization of these different effects is important for the best use of antibiotics (antibiotic stewardship). In collaboration with Lingchong You (Duke University) and with Philippe Glaser (Institut Pasteur), we develop experimental platforms, models, and optimal model calibration methods that gives precise estimations of individual resistance and collective resilience of bacterial populations to antibiotic treatments.

M3DISIM Project-Team

4. Application Domains

4.1. Clinical applications

After several validation steps – based on clinical and experimental data – we have reached the point of having validated the heart model in a pre-clinical context where we have combined direct and inverse modeling in order to bring predictive answers on specific patient states. For example, we have demonstrated the predictive ability of our model to set up pacemaker devices for a specific patient in cardiac resynchronization therapies, see [11]. We have also used our parametric estimation procedure to provide a quantitative characterization of an infarct in a clinical experiment performed with pigs, see [3].

PARIETAL Project-Team

4. Application Domains

4.1. Cognitive neuroscience

4.1.1. *Macroscopic Functional cartography with functional Magnetic Resonance Imaging (fMRI)*

The brain as a highly structured organ, with both functional specialization and a complex network organization. While most of the knowledge historically comes from lesion studies and animal electophysiological recordings, the development of non-invasive imaging modalities, such as fMRI, has made it possible to study routinely high-level cognition in humans since the early 90's. This has opened major questions on the interplay between mind and brain , such as: How is the function of cortical territories constrained by anatomy (connectivity) ? How to assess the specificity of brain regions ? How can one characterize reliably inter-subject differences ?

4.1.2. *Analysis of brain Connectivity*

Functional connectivity is defined as the interaction structure that underlies brain function. Since the beginning of fMRI, it has been observed that remote regions sustain high correlation in their spontaneous activity, i.e. in the absence of a driving task. This means that the signals observed during resting-state define a signature of the connectivity of brain regions. The main interest of resting-state fMRI is that it provides easy-to-acquire functional markers that have recently been proved to be very powerful for population studies.

4.1.3. *Modeling of brain processes (MEG)*

While fMRI has been very useful in defining the function of regions at the mm scale, Magneto-encephalography (MEG) provides the other piece of the puzzle, namely temporal dynamics of brain activity, at the ms scale. MEG is also non-invasive. It makes it possible to keep track of precise schedule of mental operations and their interactions. It also opens the way toward a study of the rhythmic activity of the brain. On the other hand, the localization of brain activity with MEG entails the solution of a hard inverse problem.

4.1.4. *Current challenges in human neuroimaging (acquisition+analysis)*

Human neuroimaging targets two major goals: *i)* the study of neural responses involved in sensory, motor or cognitive functions, in relation to models from cognitive psychology, i.e. the identification of neurophysiological and neuroanatomical correlates of cognition; *ii)* the identification of markers in brain structure and function of neurological or psychiatric diseases. Both goals have to deal with a tension between

- the search for higher spatial [0] resolution to increase **spatial specificity** of brain signals, and clarify the nature (function and structure) of brain regions. This motivates efforts for high-field imaging and more efficient acquisitions, such as compressed sensing schemes, as well as better source localization methods from M/EEG data.

- the importance of inferring brain features with **population-level** validity, hence, contaminated with high variability within observed cohorts, which blurs the information at the population level and ultimately limits the spatial resolution of these observations.

---

[0]and to some extent, temporal, but for the sake of simplicity we focus here on spatial aspects.

Importantly, the signal-to-noise ratio (SNR) of the data remains limited due to both resolution improvements [0] and between-subject variability. Altogether, these factors have led to realize that results of neuroimaging studies were **statistically weak**, i.e. plagued with low power and leading to unreliable inference [70], and particularly so due to the typically number of subjects included in brain imaging studies (20 to 30, this number tends to increase [71]): this is at the core of the *neuroimaging reproducibility crisis*. This crisis is deeply related to a second issue, namely that only few neuroimaging datasets are publicly available, making it impossible to re-assess a posteriori the information conveyed by the data. Fortunately, the situation improves, lead by projects such as NeuroVault or OpenfMRI. A framework for integrating such datasets is however still missing.

---

[0]The SNR of the acquired signal is proportional to the voxel size, hence an improvement by a factor of 2 in image resolution along each dimension is payed by a factor of 8 in terms of SNR.

<p style="color:red; text-align:center">**XPOP Project-Team**</p>

4. Application Domains

4.1. Precision medicine and pharmacogenomics

Pharmacogenomics involves using an individual's genome to determine whether or not a particular therapy, or dose of therapy, will be effective. Indeed, people's reaction to a given drug depends on their physiological state and environmental factors, but also to their individual genetic make-up.

Precision medicine is an emerging approach for disease treatment and prevention that takes into account individual variability in genes, environment, and lifestyle for each person. While some advances in precision medicine have been made, the practice is not currently in use for most diseases.

Currently, in the traditional population approach, inter-individual variability in the reaction to drugs is modeled using covariates such as weight, age, sex, ethnic origin, etc. Genetic polymorphisms susceptible to modify pharmacokinetic or pharmacodynamic parameters are much harder to include, especially as there are millions of possible polymorphisms (and thus covariates) per patient.

The challenge is to determine which genetic covariates are associated to some PKPD parameters and/or implicated in patient responses to a given drug.

Another problem encountered is the dependence of genes, as indeed, gene expression is a highly regulated process. In cases where the explanatory variables (genomic variants) are correlated, Lasso-type methods for model selection are thwarted.

There is therefore a clear need for new methods and algorithms for the estimation, validation and selection of mixed effects models adapted to the problems of genomic medicine.

A target application of this project concerns the lung cancer.

EGFR (Epidermal Growth Factor Receptor) is a cell surface protein that binds to epidermal growth factor. We know that deregulation of the downstream signaling pathway of EGFR is involved in the development of lung cancers and several gene mutations responsible for this deregulation are known.

Our objective is to identify the variants responsible for the disruption of this pathway using a modelling approach. The data that should be available for developing such model are ERK (Extracellular signal–regulated kinases) phosphorylation time series, obtained from different genetic profiles.

The model that we aim to develop will describe the relationship between the parameters of the pathway and the genomic covariates, i.e. the genetic profile. Variants related to the pathway include: variants that modify the affinity binding of ligands to receptors, variants that modify the total amount of protein, variants that affect the catalytic site,...

4.2. Oncology

In cancer, the most dreadful event is the formation of metastases that disseminate tumor cells throughout the organism. Cutaneous melanoma is a cancer, where the primary tumor can easily be removed by surgery. However, this cancer is of poor prognosis; because melanomas metastasize often and rapidly. Many melanomas arise from excessive exposure to mutagenic UV from the sun or sunbeds. As a consequence, the mutational burden of melanomas is generally high

RAC1 encodes a small GTPase that induces cell cycle progression and migration of melanoblasts during embryonic development. Patients with the recurrent P29S mutation of RAC1 have 3-fold increased odds at having regional lymph nodes invaded at the time of diagnosis. RAC1 is unlikely to be a good therapeutic target, since a potential inhibitor that would block its catalytic activity, would also lock it into the active GTP-bound state. This project thus investigates the possibility of targeting the signaling pathway downstream of RAC1.

XPOP is mainly involved in Task 1 of the project: *Identifying deregulations and mutations of the ARP2/3 pathway in melanoma patients*.

Association of over-expression or down-regulation of each marker with poor prognosis in terms of invasion of regional lymph nodes, metastases and survival, will be examined using classical univariate and multivariate analysis. We will then develop specific statistical models for survival analysis in order to associate prognosis factors to each composition of complexes. Indeed, one has to implement the further constraint that each subunit has to be contributed by one of several paralogous subunits. An original method previously developed by XPOP has already been successfully applied to WAVE complex data in breast cancer.

The developed models will be rendered user-friendly though a dedicated Rsoftware package.

This project can represent a significant step forward in precision medicine of the cutaneous melanoma.

4.3. Hemodialysis

Hemodialysis is a process for removing waste and excess water from the blood and is used primarily as an artificial replacement for lost kidney function in people with kidney failure. Side effects caused by removing too much fluid and/or removing fluid too rapidly include low blood pressure, fatigue, chest pains, leg-cramps, nausea and headaches.

Nephrologists must therefore correctly assess the hydration status in chronic hemodialysis patients and consider fluid overload effects when prescribing dialysis, according to a new study.

The fluid overload biomarker, B-type natriuretic peptide (BNP) is an important component of managing patients with kidney disease. Indeed, it is believed that each dialysis patient will have an ideal or "dry" BNP level which will accurately and reproducibly reflect their optimal fluid status.

The objective of this study is to develop a model for the BNP and the hydratation status using individual information (age, sex, ethnicity, systolic blood pressure, BMI, coronary heart disease history, ...).

The impact will be significant if the method succeeds. Indeed, it will be possible for the nephrologists to use this model for monitoring individually each treatment, in order to avoid risks of hypotension (low BNP) or overweight (high BNP).

4.4. Intracellular processes

Significant cell-to-cell heterogeneity is ubiquitously-observed in isogenic cell populations. Cells respond differently to a same stimulation. For example, accounting for such heterogeneity is essential to quantitatively understand why some bacteria survive antibiotic treatments, some cancer cells escape drug-induced suicide, stem cell do not differentiate, or some cells are not infected by pathogens.

The origins of the variability of biological processes and phenotypes are multifarious. Indeed, the observed heterogeneity of cell responses to a common stimulus can originate from differences in cell phenotypes (age, cell size, ribosome and transcription factor concentrations, etc), from spatio-temporal variations of the cell environments and from the intrinsic randomness of biochemical reactions. From systems and synthetic biology perspectives, understanding the exact contributions of these different sources of heterogeneity on the variability of cell responses is a central question.

The main ambition of this project is to propose a paradigm change in the quantitative modelling of cellular processes by shifting from mean-cell models to single-cell and population models. The main contribution of XPOP focuses on methodological developments for mixed-effects model identification in the context of growing cell populations.

- Mixed-effects models usually consider an homogeneous population of independent individuals. This assumption does not hold when the population of cells (i.e. the statistical individuals) consists of several generations of dividing cells. We then need to account for inheritance of single-cell parameters in this population. More precisely, the problem is to attribute the new state and parameter values to newborn cells given (the current estimated values for) the mother.

- The mixed-effects modelling framework corresponds to a strong assumption: differences between cells are static in time (ie, cell-specific parameters have fixed values). However, it is likely that for any given cell, ribosome levels slowly vary across time, since like any other protein, ribosomes are produced in a stochastic manner. We will therefore extend our modelling framework so as to account for the possible random fluctuations of parameter values in individual cells. Extensions based on stochastic differential equations will be investigated.

- Identifiability is a fundamental prerequisite for model identification and is also closely connected to optimal experimental design. We will derive criteria for theoretical identifiability, in which different parameter values lead to non-identical probability distributions, and for structural identifiability, which concerns the algebraic properties of the structural model, i.e. the ODE system. We will then address the problem of practical identifiability, whereby the model may be theoretically identifiable but the design of the experiment may make parameter estimation difficult and imprecise. An interesting problem is whether accounting for lineage effects can help practical identifiability of the parameters of the individuals in presence of measurement and biological noise.

4.5. Population pharmacometrics

Pharmacometrics involves the analysis and interpretation of data produced in pre-clinical and clinical trials. Population pharmacokinetics studies the variability in drug exposure for clinically safe and effective doses by focusing on identification of patient characteristics which significantly affect or are highly correlated with this variability. Disease progress modeling uses mathematical models to describe, explain, investigate and predict the changes in disease status as a function of time. A disease progress model incorporates functions describing natural disease progression and drug action.

The model based drug development (MBDD) approach establishes quantitative targets for each development step and optimizes the design of each study to meet the target. Optimizing study design requires simulations, which in turn require models. In order to arrive at a meaningful design, mechanisms need to be understood and correctly represented in the mathematical model. Furthermore, the model has to be predictive for future studies. This requirement precludes all purely empirical modeling; instead, models have to be mechanistic.

In particular, physiologically based pharmacokinetic models attempt to mathematically transcribe anatomical, physiological, physical, and chemical descriptions of phenomena involved in the ADME (Absorption - Distribution - Metabolism - Elimination) processes. A system of ordinary differential equations for the quantity of substance in each compartment involves parameters representing blood flow, pulmonary ventilation rate, organ volume, etc.

The ability to describe variability in pharmacometrics model is essential. The nonlinear mixed-effects modeling approach does this by combining the structural model component (the ODE system) with a statistical model, describing the distribution of the parameters between subjects and within subjects, as well as quantifying the unexplained or residual variability within subjects.

The objective of XPOP is to develop new methods for models defined by a very large ODE system, a large number of parameters and a large number of covariates. Contributions of XPOP in this domain are mainly methodological and there is no privileged therapeutic application at this stage.

However, it is expected that these new methods will be implemented in software tools, including MONOLIX and Rpackages for practical use.

4.6. Mass spectrometry

One of the main recent developments in analytical chemistry is the rapid democratization of high-resolution mass spectrometers. These instruments produce extremely complex mass spectra, which can include several hundred thousand ions when analyzing complex samples. The analysis of complex matrices (biological, agri-food, cosmetic, pharmaceutical, environmental, etc.) is precisely one of the major analytical challenges of this new century. Academic and industrial researchers are particularly interested in trying to quickly and effectively establish the chemical consequences of an event on a complex matrix. This may include, for

example, searching for pesticide degradation products and metabolites in fruits and vegetables, photoproducts of active ingredients in a cosmetic emulsion exposed to UV rays or chlorination products of biocides in hospital effluents. The main difficulty of this type of analysis is based on the high spatial and temporal variability of the samples, which is in addition to the experimental uncertainties inherent in any measurement and requires a large number of samples and analyses to be carried out and computerized data processing (up to 16 million per mass spectrum).

A collaboration between XPOP and the Molecular Chemistry Laboratory (LCM) of the Ecole Polytechnique began in 2018. Our objective is to develop new methods for the statistical analysis of mass spectrometry data.

INFINE-POST Team (section vide)

POEMS-POST Team

4. Application Domains

4.1. Acoustics

Two particular subjects have retained our attention recently.

1. Aeroacoustics, or more precisely, acoustic propagation in a moving compressible fluid, has been for our team a very challenging topic, which gave rise to a lot of open questions, from the modeling until the numerical approximation of existing models. Our works in this area are partially supported by Airbus. The final objective is to reduce the noise radiated by Airbus planes.

2. Musical acoustics constitute a particularly attractive application. We are concerned by the simulation of musical instruments whose objectives are both a better understanding of the behavior of existing instruments and an aid for the manufacturing of new instruments. We have successively considered the timpani, the guitar and the piano.

4.2. Electromagnetism

Applied mathematics for electromagnetism during the last ten years have mainly concerned stealth technology and electromagnetic compatibility. These areas are still motivating research in computational sciences (large scale computation) and mathematical modeling (derivation of simplified models for multiscale problems). Electromagnetic propagation in non classical media opens a wide and unexplored field of research in applied mathematics. This is the case of wave propagation in photonic crystals, metamaterials or magnetized plasmas. Finally, the simulation electromagnetic (possibly complex, even fractal) networks is motivated by destructive testing applications. These topics are developed in collaboration with CEA, DGA and ONERA.

4.3. Elastodynamics

Wave propagation in solids is with no doubt, among the three fundamental domains that are acoustics, electromagnetism and elastodynamics, the one that poses the most significant difficulties from mathematical and numerical points of view. A major application topic has emerged during the past years : the non destructive testing by ultra-sounds which is the main topic of our collaboration with CEA-LIST. On the other hand, we are developing efficient integral equation modelling for geophysical applications (soil-structure interaction for civil engineering, seismology).

AVIZ Project-Team (section vide)

CEDAR Project-Team

4. Application Domains

4.1. Cloud Computing

Cloud computing services are strongly developing and more and more companies and institutions resort to running their computations in the cloud, in order to avoid the hassle of running their own infrastructure. Today's cloud service providers guarantee machine availabilities in their Service Level Agreement (SLA), without any guarantees on performance measures according to a specific cost budget. Running analytics on big data systems require the user not to only reserve the suitable cloud instances over which the big data system will be running, but also setting many system parameters like the degree of parallelism and granularity of scheduling. Chosing values for these parameters, and chosing cloud instances need to meet user objectives regarding latency, throughput and cost measures, which is a complex task if it's done manually by the user. Hence, we need need to transform cloud service models from availabily to user performance objective rises and leads to the problem of multi-objective optimization. Research carried out in the team within the ERC project "Big and Fast Data Analytics" aims to develop a novel optimization framework for providing guarantees on the performance while controlling the cost of data processing in the cloud.

4.2. Computational Journalism

Modern journalism increasingly relies on content management technologies in order to represent, store, and query source data and media objects themselves. Writing news articles increasingly requires consulting several sources, interpreting their findings in context, and crossing links between related sources of information. CEDARresearch results directly applicable to this area provide techniques and tools for rich Web content warehouse management. Within the ANR ContentCheck project, and also as part of our international collaboration with the AIST institute from Japan, we work on one hand, to lay down foundations for computational data journalism and fact checking, and also work to devise concrete algorithms and platforms to help journalists perform their work better and/or faster. This work is carried in collaboration with Le Monde's "Les Décodeurs".

On a related topic, heterogeneous data integration under a virtual graph abstract model is studied within the ICODA Inria project which has started in September 2017. There, we collaborate with Les Décodeurs as well as with Ouest France and Agence France Presse (AFP). The data and knowledge integration framework resulting from this work will support journalists' effort to organize and analyze their knowledge and exploit it in order to produce new content.

4.3. Open Data Intelligence

The Web is a vast source of information, to which more is added every day either in unstructured form (Web pages) or, increasingly, as partially structured sources of information, in particular as Open Data sets, which can be seen as connected graphs of data, most frequently described in the RDF data format recommended by the W3C. Further, RDF data is also the most appropriate format for representing structured information extracted automatically from Web pages, such as the DBPedia database extracted from Wikipedia or Google's InfoBoxes. Our work on this topic has taken place within the 4-year project ODIN, funded by the Department of Defense under the RAPID innovation programme.

4.4. Genomics

One particular case of area where the increase in data production is the more consequent is genomic data, indeed the amount of data produced doubles every 7 months. Thus we want to bring the expertise from the database and big data community to help both scale the existing algorithms and design new algorithms that are scalable from the ground up.

EX-SITU Project-Team

4. Application Domains

4.1. Creative industries

We work closely with creative professionals in the arts and in design, including music composers, musicians, and sound engineers; painters and illustrators; dancers and choreographers; theater groups; game designers; graphic and industrial designers; and architects.

4.2. Scientific research

We work with creative professionals in the sciences and engineering, including neuroscientists and doctors; programmers and statisticians; chemists and astrophysicists; and researchers in fluid mechanics.

ILDA Project-Team

4. Application Domains

4.1. Mission-critical systems

Mission-critical contexts of use include emergency response & management, and critical infrastructure operations, such as public transportation systems, communications and power distribution networks, or the operations of large scientific instruments such as particle accelerators and astronomical observatories. Central to these contexts of work is the notion of situation awareness [33], i.e., how workers perceive and understand elements of the environment with respect to time and space, such as maps and geolocated data feeds from the field, and how they form mental models that help them predict future states of those elements. One of the main challenges is how to best assist subject-matter experts in constructing correct mental models and making informed decisions, often under time pressure. This can be achieved by providing them with, or helping them efficiently identify and correlate, relevant and timely information extracted from large amounts of raw data, taking into account the often cooperative nature of their work and the need for task coordination. With this application area, our goal is to investigate novel ways of interacting with computing systems that improve collaborative data analysis capabilities and decision support assistance in a mission-critical, often time-constrained, work context.

Relevant publications by team members this year: [25], [31], [16], [21], [23], [28], [13].

4.2. Exploratory analysis of scientific data

Many scientific disciplines are increasingly data-driven, including astronomy, molecular biology, particle physics, or neuroanatomy. While making the right decision under time pressure is often less of critical issue when analyzing scientific data, at least not on the same temporal scale as truly time-critical systems, scientists are still faced with large-to-huge amounts of data. No matter their origin (experiments, remote observations, large-scale simulations), these data are difficult to understand and analyze in depth because of their sheer size and complexity. Challenges include how to help scientists freely-yet-efficiently explore their data, keep a trace of the multiple data processing paths they considered to verify their hypotheses and make it easy to backtrack, and how to relate observations made on different parts of the data and insights gained at different moments during the exploration process. With this application area, our goal is to investigate how data-centric interactive systems can improve collaborative scientific data exploration, where users' goals are more open-ended, and where roles, collaboration and coordination patterns [53] differ from those observed in mission-critical contexts of work.

Relevant publications by team members last year: [15], [23], [30].

PETRUS Project-Team

4. Application Domains

4.1. Personal cloud, home care, IoT, sensing, surveys

As stated in the software section, the Petrus research strategy aims at materializing its scientific contributions in an advanced hardware/software platform with the expectation to produce a real societal impact. Hence, our software activity is structured around a common Secure Personal Cloud platform rather than several isolated demonstrators. This platform will serve as the foundation to develop a few emblematic applications. Several privacy-preserving applications can actually be targeted by a Personal Cloud platform, like: (i) smart disclosure applications allowing the individual to recover her personal data from external sources (e.g., bank, online shopping activity, insurance, etc.), integrate them and cross them to perform personal big data tasks (e.g., to improve her budget management) ; (ii) management of personal medical records for care coordination and well-being improvement; (iii) privacy-aware data management for the IoT (e.g., in sensors, quantified-self devices, smart meters); (iv) community-based sensing and community data sharing; (v) privacy-preserving studies (e.g., cohorts, public surveys, privacy-preserving data publishing). Such applications overlap with all the research axes described above but each of them also presents its own specificities. For instance, the smart disclosure applications will focus primarily on sharing models and enforcement, the IoT applications require to look with priority at the embedded data management and sustainability issues, while community-based sensing and privacy-preserving studies demand to study secure and efficient global query processing. Among these applications domains, one is already receiving a particular attention from our team. Indeed, we gained a strong expertise in the management and protection of healthcare data through our past DMSP (Dossier Medico-Social Partagé) experiment in the field. This expertise is being exploited to develop a dedicated healthcare and well-being personal cloud platform. We are currently deploying 10000 boxes equipped with PlugDB in the context of the DomYcile project. In this context, we are currently setting up an Inria Innovation Lab with the Hippocad company to industrialize this platform and deploy it at large scale (see Section the bilateral contract OwnCare II-Lab).