RESEARCH CENTER

FIELD
Algorithmics, Programming, Software and Architecture

Activity Report 2019

Section Application Domains

Edition: 2020-03-21

SECURITY AND CONFIDENTIALITY

ARIC Project-Team

4. Application Domains

4.1. Floating-point and Validated Numerics

Our expertise on validated numerics is useful to analyze and improve, and guarantee the quality of numerical results in a wide range of applications including:

- scientific simulation;
- global optimization;
- control theory.

Much of our work, in particular the development of correctly rounded elementary functions, is critical to the

- reproducibility of floating-point computations.

4.2. Cryptography, Cryptology, Communication Theory

Lattice reduction algorithms have direct applications in

- public-key cryptography;
- diophantine equations;
- communications theory.

AROMATH Project-Team

4. Application Domains

4.1. Geometric modeling for Design and Manufacturing.

The main domain of applications that we consider for the methods we develop is Computer Aided Design and Manufacturing.

Computer-Aided Design (CAD) involves creating digital models defined by mathematical constructions, from geometric, functional or aesthetic considerations. Computer-aided manufacturing (CAM) uses the geometrical design data to control the tools and processes, which lead to the production of real objects from their numerical descriptions.

CAD-CAM systems provide tools for visualizing, understanding, manipulating, and editing virtual shapes. They are extensively used in many applications, including automotive, shipbuilding, aerospace industries, industrial and architectural design, prosthetics, and many more. They are also widely used to produce computer animation for special effects in movies, advertising and technical manuals, or for digital content creation. Their economic importance is enormous. Their importance in education is also growing, as they are more and more used in schools and educational purposes.

CAD-CAM has been a major driving force for research developments in geometric modeling, which leads to very large software, produced and sold by big companies, capable of assisting engineers in all the steps from design to manufacturing.

Nevertheless, many challenges still need to be addressed. Many problems remain open, related to the use of efficient shape representations, of geometric models specific to some application domains, such as in architecture, naval engineering, mechanical constructions, manufacturing ...Important questions on the robustness and the certification of geometric computation are not yet answered. The complexity of the models which are used nowadays also appeals for the development of new approaches. The manufacturing environment is also increasingly complex, with new type of machine tools including: turning, 5-axes machining and wire EDM (Electrical Discharge Machining), 3D printer. It cannot be properly used without computer assistance, which raises methodological and algorithmic questions. There is an increasing need to combine design and simulation, for analyzing the physical behavior of a model and for optimal design.

The field has deeply changed over the last decades, with the emergence of new geometric modeling tools built on dedicated packages, which are mixing different scientific areas to address specific applications. It is providing new opportunities to apply new geometric modeling methods, output from research activities.

4.2. Geometric modeling for Numerical Simulation and Optimization

A major bottleneck in the CAD-CAM developments is the lack of interoperability of modeling systems and simulation systems. This is strongly influenced by their development history, as they have been following different paths.

The geometric tools have evolved from supporting a limited number of tasks at separate stages in product development and manufacturing, to being essential in all phases from initial design through manufacturing.

Current Finite Element Analysis (FEA) technology was already well established 40 years ago, when CAD-systems just started to appear, and its success stems from using approximations of both the geometry and the analysis model with low order finite elements (most often of degree ≤ 2).

There has been no requirement between CAD and numerical simulation, based on Finite Element Analysis, leading to incompatible mathematical representations in CAD and FEA. This incompatibility makes interoperability of CAD/CAM and FEA very challenging. In the general case today this challenge is addressed by expensive and time-consuming human intervention and software developments.

Improving this interaction by using adequate geometric and functional descriptions should boost the interaction between numerical analysis and geometric modeling, with important implications in shape optimization. In particular, it could provide a better feedback of numerical simulations on the geometric model in a design optimization loop, which incorporates iterative analysis steps.

The situation is evolving. In the past decade, a new paradigm has emerged to replace the traditional Finite Elements by B-Spline basis element of any polynomial degree, thus in principle enabling exact representation of all shapes that can be modeled in CAD. It has been demonstrated that the so-called isogeometric analysis approach can be far more accurate than traditional FEA.

It opens new perspectives for the interoperability between geometric modeling and numerical simulation. The development of numerical methods of high order using a precise description of the shapes raises questions on piecewise polynomial elements, on the description of computational domains and of their interfaces, on the construction of good function spaces to approximate physical solutions. All these problems involve geometric considerations and are closely related to the theory of splines and to the geometric methods we are investigating. We plan to apply our work to the development of new interactions between geometric modeling and numerical solvers.

CARAMBA Project-Team

4. Application Domains

4.1. Better Awareness and Avoidance of Cryptanalytic Threats

Our study of the Number Field Sieve family of algorithms aims at showing how the threats underlying various supposedly hard problems are real. Our record computations, as well as new algorithms, contribute to having a scientifically accurate assessment of the feasibility limit for these problems, given academic computing resources. The data we provide in this way is a primary ingredient for government agencies whose purpose includes guidance for the choice of appropriate cryptographic primitives. For example the French ANSSI [0], German BSI, or the NIST [0] in the United States base their recommendations on such computational achievements.

The software we make available to achieve these cryptanalytic computations also allows us to give cost estimates for potential attacks to cryptographic systems that are taking the security/efficiency/legacy compatibility trade-offs too lightly. Attacks such as LogJam [26] are understood as being serious concerns thanks to our convincing proof-of-concepts. In the LogJam context, this impact has led to rapid worldwide security advisories and software updates that eventually defeat some potential intelligence threats and improve confidentiality of communications.

4.2. Promotion of Better Cryptography

We also promote the switch to algebraic curves as cryptographic primitives. Those offer nice speed and excellent security, while primitives based on elementary number theory (integer factorization, discrete logarithm in finite fields), which underpin e.g., RSA, are gradually forced to adopt unwieldy key sizes so as to comply with the desired security guarantees of modern cryptography. Our contributions to the ultimate goal of having algebraic curves eventually take over the cryptographic landscape lie in our fast arithmetic contributions, our contributions to the point counting problem, and more generally our expertise on the diverse surrounding mathematical objects, or on the special cases where the discrete logarithm problem is not hard enough and should be avoided.

We also promote cryptographically sound electronic voting, for which we develop the Belenios prototype software (licensed under the AGPL). It depends on research made in collaboration with the PESTO team, and provides stronger guarantees than current state of the art.

4.3. Key Software Tools

The vast majority of our work is eventually realized as software. We can roughly categorize it in two groups. Some of our software covers truly fundamental objects, such as the GNU MPFR, GNU MPC, GF2X, or MPFQ packages. To their respective extent, these software packages are meant to be included or used in broader projects. For this reason, it is important that the license chosen for this software allows proper reuse, and we favor licenses such as the LGPL, which is not restrictive. We can measure the impact of this software by the way it is used in e.g., the GNU Compiler Collection (GCC), in Victor Shoup's Number Theory Library (NTL), or in the Sage computer algebra system. The availability of these software packages in most Linux distributions is also a good measure for the impact of our work.

[0] In [27], the minimal recommended RSA key size is 2048 bits for usage up to 2030. See also Annex B, in particular Section B.1 "Records de calculs cryptographiques".

[0] The work [31] is one of only two academic works cited by NIST in the initial version (2011) of the report [33].

We also develop more specialized software. Our flagship software package is Cado-NFS [15], and we also develop some others with various levels of maturity, such as GMP-ECM, CMH, or Belenios, aiming at quite diverse targets. Within the lifespan of the CARAMBA project, we expect more software packages of this kind to be developed, specialized towards tasks relevant to our research targets: important mathematical structures attached to genus 2 curves, generation of cryptographically secure curves, or tools for attacking cryptographically hard problems. Such software both illustrates our algorithms, and provides a base on which further research work can be established. Because of the very nature of these specialized software packages as research topics in their own right, needing both to borrow material from other projects, and being possible source of inspiring material for others, it is again important that these be developed in a free and open-source development model.

CASCADE Project-Team

4. Application Domains

4.1. Privacy for the Cloud

Many companies have already started the migration to the Cloud and many individuals share their personal informations on social networks. While some of the data are public information, many of them are personal and even quite sensitive. Unfortunately, the current access mode is purely right-based: the provider first authenticates the client, and grants him access, or not, according to his rights in the access-control list. Therefore, the provider itself not only has total access to the data, but also knows which data are accessed, by whom, and how: privacy, which includes secrecy of data (confidentiality), identities (anonymity), and requests (obliviousness), should be enforced. Moreover, while high availability can easily be controlled, and thus any defect can immediately be detected, failures in privacy protection can remain hidden for a long time. The industry of the Cloud introduces a new implicit trust requirement: nobody has any idea at all of where and how his data are stored and manipulated, but everybody should blindly trust the providers. The providers will definitely do their best, but this is not enough. Privacy-compliant procedures cannot be left to the responsibility of the provider: however strong the trustfulness of the provider may be, any system or human vulnerability can be exploited against privacy. This presents too huge a threat to tolerate. *The distribution of the data and the secrecy of the actions must be given back to the users. It requires promoting privacy as a global security notion.*

In order to protect the data, one needs to encrypt it. Unfortunately, traditional encryption systems are inadequate for most applications involving big, complex data. Recall that in traditional public key encryption, a party encrypts data to a single known user, which lacks the expressiveness needed for more advanced data sharing. In enterprise settings, a party will want to share data with groups of users based on their credentials. Similarly, individuals want to selectively grant access to their personal data on social networks as well as documents and spreadsheets on Google Docs. Moreover, the access policy may even refer to users who do not exist in the system at the time the data is encrypted. Solving this problem requires an entirely new way of encrypting data.

A first natural approach would be **fully homomorphic encryption** (FHE, see above), but a second one is also **functional encryption**, that is an emerging paradigm for public-key encryption: it enables more fine-grained access control to encrypted data, for instance, the ability to specify a decryption policy in the ciphertext so that only individuals who satisfy the policy can decrypt, or the ability to associate keywords to a secret key so that it can only decrypt documents containing the keyword. Our work on functional encryption centers around two goals:

1. to obtain more efficient pairings-based functional encryption;
2. and to realize new functionalities and more expressive functional encryption schemes.

Another approach is **secure multi-party computation protocols**, where interactivity might provide privacy in a more efficient way. Recent implicit interactive proofs of knowledge can be a starting point. But stronger properties are first expected for improving privacy. They can also be integrated into new ad-hoc broadcast systems, in order to distribute the management among several parties, and eventually remove any trust requirements.

Strong privacy for the Cloud would have a huge societal impact since it would revolutionize the trust model: users would be able to make safe use of outsourced storage, namely for personal, financial and medical data, without having to worry about failures or attacks of the server.

DATASHAPE Project-Team (section vide)

GAMBLE Project-Team

4. Application Domains

4.1. Applications of computational geometry

Many domains of science can benefit from the results developed by GAMBLE. Curves and surfaces are ubiquitous in all sciences to understand and interpret raw data as well as experimental results. Still, the non-linear problems we address are rather basic and fundamental, and it is often difficult to predict the impact of solutions in that area. The short-term industrial impact is likely to be small because, on basic problems, industries have used ad hoc solutions for decades and have thus got used to it. The example of our work on quadric intersection is typical: even though we were fully convinced that intersecting 3D quadrics is such an elementary/fundamental problem that it ought to be useful, we were the first to be astonished by the scope of the applications of our software [0] (which was the first and still is the only one —to our knowledge— to compute robustly and efficiently the intersection of 3D quadrics) which has been used by researchers in, for instance, photochemistry, computer vision, statistics, and mathematics. Our work on certified drawing of plane (algebraic) curves falls in the same category. It seems obvious that it is widely useful to be able to draw curves correctly (recall also that part of the problem is to determine where to look in the plane) but it is quite hard to come up with specific examples of fields where this is relevant. A contrario, we know that certified meshing is critical in mechanical-design applications in robotics, which is a non-obvious application field. There, the singularities of a manipulator often have degrees higher than 10 and meshing the singular locus in a certified way is currently out of reach. As a result, researchers in robotics can only build physical prototypes for validating, or not, the approximate solutions given by non-certified numerical algorithms.

The fact that several of our pieces of software for computing non-Euclidean triangulations had already been requested by users long before they become public in CGAL is a good sign for their wide future impact. This will not come as a surprise, since most of the questions that we have been studying followed from discussions with researchers outside computer science and pure mathematics. Such researchers are either users of our algorithms and software, or we meet them in workshops. Let us only mention a few names here. Rien van de Weijgaert [55], [69] (astrophysicist, Groningen, NL) and Michael Schindler [66] (theoretical physicist, ENSPCI, CNRS, France) used our software for 3D periodic weighted triangulations. Stephen Hyde and Vanessa Robins (applied mathematics and physics at Australian National University) used our package for 3D periodic meshing. Olivier Faugeras (neuromathematics, Inria Sophia Antipolis) had come to us and mentioned his needs for good meshes of the Bolza surface [45] before we started to study them. Such contacts are very important both to get feedback about our research and to help us choose problems that are relevant for applications. These problems are at the same time challenging from the mathematical and algorithmic points of view. Note that our research and our software are generic, i.e., we are studying fundamental geometric questions, which do not depend on any specific application. This recipe has made the sucess of the CGAL library.

Probabilistic models for geometric data are widely used to model various situations ranging from cell phone distribution to quantum mechanics. The impact of our work on probabilistic distributions is twofold. On the one hand, our studies of properties of geometric objects built on such distributions will yield a better understanding of the above phenomena and has potential impact in many scientific domains. On the other hand, our work on simulations of probabilistic distributions will be used by other teams, more maths oriented, to study these distributions.

[0] QI: web.

GRACE Project-Team

4. Application Domains

4.1. Application Domain: cybersecurity

Participants: Guénaël Renault, Benjamin Smith, François Morain, Alexis Challande, Simon Montoya, Maxime Anvari.

We are interesting in developing some interactions between cryptography and cybersecurity. In particular, we develop some researches in embedded security (side channels and fault attack), software security (finding vulnerability efficiently) and privacy (security of TOR).

4.2. Application Domain: blockchains

Participants: Daniel Augot, Sarah Bordage, Matthieu Rambaud, Lucas Benmouffok, Hanna-Mae Bisserier.

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.

Daniel Augot is involved in blockchains from the point of view of cryptography for better blockchains, mainly for improving privacy. A PhD student has been enrolled at IRT System-X, to study pratical use cases of Secure Multiparty Computaiton.

Also Daniel Augot, together with Julian Prat (economist, ENSAE), is leading a Polytechnique teaching and research “chair”, funded by CapGemini, for blockchains in the industry, B2B platforms, supply chains, etc.

4.3. 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).

An M2 intern, Maxime Roméas, Bordeaux university, studied the constructive cryptography model, "A study of the Constructive Cryptography model of Maurer et. al." 5 months, followed by a PhD grant from IP Paris/Ecole Polytechnique for a 3-year doctorate (Oct 2019-Sept 2022): "The Constructive Cryptography paradigm applied to Interactive Cryptographic Proofs".

The Constructive Cryptography framework redefines basic cryptographic primitives and protocols starting from discrete systems of three types (resources, converters, and distinguishers). This not only permits to construct them effectively, but also lighten and sharpen their security proofs. One strength of this model is its composability. The purpose of the PhD is to apply this model to rephrase existing interactive cryptographic proofs so as to assert their genuine security, as well as to design new proofs. The main concern here is security and privacy in Distributed Storage settings.

LFANT Project-Team (section vide)

OURAGAN Project-Team

4. Application Domains

4.1. Security of cryptographic systems

The study of the security of asymmetric cryptographic systems comes as an application of the work carried out in algorithmic number theory and revolves around the development and the use of a small number of general purpose algorithms (lattice reduction, class groups in number fields, discrete logarithms in finite fields, ...). For example, the computation of generators of principal ideals of cyclotomic fields can be seen as one of these applications since these are used in a number of recent public key cryptosystems.

The cryptographic community is currently very actively assessing the threat coming for the development of quantum computers. Indeed, such computers would permit tremendous progress on many number theoretic problems such as factoring or discrete logarithm computations and would put the security of current cryptosystem under a major risk. For this reason, there is a large global research effort dedicated to finding alternative methods of securing data. For example, the US standardization agency called NIST has recently launched a standardization process around this issue. In this context, OURAGAN is part of the competition and has submitted a candidate (which has not been selected) [40]. This method is based on number-theoretic ideas involving a new presumably difficult problem concerning the Hamming distance of integers modulo large numbers of Mersenne.

4.2. Robotics

Algebraic computations have tremendously been used in Robotics, especially in kinematics, since the last quarter of the 20th century [93]. For example, one can find algebraic proofs for the 40 possible solutions to the direct kinematics problem [117] for steward platforms and companion experiments based on Gröbner basis computations [83]. On the one hand, hard general kinematics problems involve too many variables for pure algebraic methods to be used in place of existing numerical or semi-numerical methods everywhere and everytime, and on the other hand, global algebraic studies allow to propose exhaustive classifications that cannot be reached by other methods,for some quite large classes.

Robotics is a long-standing collaborative work with LS2N (Laboratory of Numerical Sciences of Nantes). Work has recently focused on the offline study of mechanisms, mostly parallel, their singularities or at least some types of singularities (cuspidals robots [140]).

For most parallel or serial manipulators, pose variables and joints variables are linked by algebraic equations and thus lie an algebraic variety. The two-kinematics problems (the direct kinematics problem - DKP- and the inverse kinematics problem - IKP) consist in studying the preimage of the projection of this algebraic variety onto a subset of unknowns. Solving the DKP remains to computing the possible positions for a given set of joint variables values while solving the IKP remains to computing the possible joints variables values for a given position. Algebraic methods have been deeply used in several situations for studying parallel and serial mechanisms, but finally their use stays quite confidential in the design process. Cylindrical Algebraic Decomposition coupled with variable's eliminations by means of Gröbner based computations can be used to model the workspace, the joint space and the computation of singularities. On the one hand, such methods suffer immediately when increasing the number of parameters or when working with imprecise data. On the other hand, when the problem can be handled, they might provide full and exhaustive classifications. The tools we use in that context [60], [59], [95], [97], [96] depend mainly on the resolution of parameter-based systems and therefore of study-dependent curves or flat algebraic surfaces (2 or 3 parameters), thus joining our thematic *Computational Geometry*.

4.3. Control theory

Certain problems studied in mathematical systems theory and control theory can be better understood and finely studied by means of algebraic structures and methods. Hence, the rich interplay between algebra, computer algebra, and control theory has a long history. For instance, the first main paper on Gröbner bases written by their creators, Buchberger, was published in Bose's book [46] on control theory of multidimensional systems. Moreover, the differential algebra approach to nonlinear control theory (see [72], [73] and the references therein) was a major motivation for the algorithmic study of differential algebra [47], [76]. Finally, the behaviour approach to linear systems theory [141], [119] advocates for an algorithmic study of algebraic analysis (see Section 2.2.4). More generally, control theory is porous to computer algebra since one finds algebraic criteria of all kinds in the literature even if the control theory community has a very few knowledge in computer algebra.

OURAGAN has a strong interest in the computer algebra aspects of mathematical systems theory and control theory related to both functional and polynomial systems, particularly in the direction of robust stability analysis and robust stabilization problems for multidimensional systems [46], [119] and infinite-dimensional systems [66] (such as, e.g., differential time-delay systems).

Let us shortly state a few points of our recent interests in this direction.

In control theory, stability analysis of linear time-invariant control systems is based on the famous Routh-Hurwitz criterion (late 19th century) and its relation with Sturm sequences and Cauchy index. Thus, stability tests were only involving tools for univariate polynomials [102]. While extending those tests to multidimensional systems or differential time-delay systems, one had to tackle multivariate problems recursively with respect to the variables [46]. Recent works use a mix of symbolic/numeric strategies, Linear Matrix Inequalities (LMI), sums of squares, etc. But still very few practical experiments are currently involving certified algebraic computations based on general solvers for polynomial equations. We have recently started to study certified stability tests for multidimensional systems or differential time-delay systems with an important observation: with a correct modelization, some recent algebraic methods − derived from our work in algorithmic geometry and shared with applications in robotics − can now handle previously impossible computations and lead to a better understanding of the problems to be solved [52], [54], [55]. The previous approaches seem to be blocked on a recursive use of one-variable methods, whereas our approach involves the direct processing of the problem for a larger number of variables.

The structural stability of n-D discrete linear systems (with $n \geq 2$) is a good source of problems of several kinds ranging from solving univariate polynomials to studying algebraic systems depending on parameters. For instance, we show [53], [54], [55] that the standard characterization of the structural stability of a multivariate rational transfer function (namely, the denominator of the transfer function does not have solutions in the unit polydisc of $\mathbb{C}^n$) is equivalent to deciding whether or not a certain system of polynomial equations has real solutions. The use state-of-the-art computer algebra algorithms to check this last condition, and thus the structural stability of multidimensional systems has been validated in several situations from toy examples with parameters to state-of-the-art examples involving, e.g., the resolution of bivariate systems [51], [50].

The rich interplay between control theory, algebra, and computer algebra is also well illustrated with our recent work on robust stabilization problems for multidimensional and finite/infinite-dimensional systems [48], [123], [129], [132], [130], [131].

4.4. Signal processing

Due to numerous applications (e.g. sensor network, mobile robots), sources and sensors localization has intensively been studied in the literature of signal processing. The *anchor position self calibration problem* is a well-known problem which consists in estimating the positions of both the moving sources and a set of fixed sensors (anchors) when only the distance information between the points from the different sets is available. The position self-calibration problem is a particular case of the *Multidimensional Unfolding* (MDU) problem for the Euclidean space of dimension 3. In the signal processing literature, this problem is attacked by means of optimization problems (see [65] and the references therein). Based on computer algebra methods for

polynomial systems, we have recently developed a new approach for the MDU problem which yields closed-form solutions and a very efficient algorithm for the estimation of the positions [68] based only on linear algebra techniques. This first result, done in collaboration with Dagher (Inria Chile) and Zheng (DEFROST, Inria Lille), yielded a recent patent [67]. This result advocates for the study of other localization problems based on the computational polynomial techniques developed in OURAGAN.

In collaboration with *Safran Tech* (Barau, Hubert) and Dagher (Inria Chile), a symbolic-numeric study of the new *multi-carrier demodulation method* [92] has recently been initiated. *Gear fault diagnosis* is an important issue in aeronautics industry since a damage in a gearbox, which is not detected in time, can have dramatic effects on the safety of a plane. Since the vibrations of a spur gear can be modeled as a product of two periodic functions related to the gearbox kinematic, it is proposed to recover each function from the global signal by means of an optimal reconstruction problem which, based on Fourier analysis, can be rewritten as $\operatorname{argmin}_{u\in\mathbb{C}^n, v_1, v_2\in\mathbb{C}^m} \|M - u\, v_1^{\bigstar} - D\, u\, v_2^{\bigstar}\|_F$, where $M \in \mathbb{C}^{n\times m}$ (resp. $D \in \mathbb{C}^{n\times n}$) is a given matrix with a special shape (resp. diagonal matrix), $\|\cdot\|_F$ is the Frobenius norm, and $v^{\bigstar}$ is the Hermitian transpose of v. We have recently obtained closed-form solutions for the exact problem, i.e., $M = u\, v_1^{\bigstar} + D\, u\, v_2^{\bigstar}$, which is a polynomial system with parameters. This first result gives interesting new insides for the study of the non-exact case, i.e. for the above optimization problem.

Our expertise on *algebraic parameter estimation problem*, developed in the former NON-A project-team (Inria Lille), will be further developed. Following this work [84], the problem consists in estimating a set θ of parameters of a signal $x(\theta, t)$– which satisfies a certain dynamics – when the signal $y(t) = x(\theta, t) + \gamma(t) + \varpi(t)$ is observed, where γ denotes a structured perturbation and ϖ a noise. It has been shown that θ can sometimes be explicitly determined by means of closed-form expressions using iterated integrals of y. These integrals are used to filter the noise ϖ. Based on a combination of algebraic analysis techniques (rings of differential operators), differential elimination theory (Gröbner basis techniques for Weyl algebras), and operational calculus (Laplace transform, convolution), an algorithmic approach to algebraic parameter estimation problem has been initiated in [125] for a particular type of structured perturbations (i.e. bias) and was implemented in the `Maple` prototype `NonA`. The case of a general structured perturbation is still lacking.

POLSYS Project-Team (section vide)

SECRET Project-Team

4. Application Domains

4.1. Cryptographic primitives

Our major application domain is the design of cryptographic primitives, especially for platforms with restricting implementation requirements. For instance, we aim at recommending (or designing) low-cost (or extremely fast) encryption schemes, or primitives which remain secure against quantum computers.

4.2. Code Reconstruction

To evaluate the quality of a cryptographic algorithm, it is usually assumed that its specifications are public, as, in accordance with Kerckhoffs principle, it would be dangerous to rely, even partially, on the fact that the adversary does not know those specifications. However, this fundamental rule does not mean that the specifications are known to the attacker. In practice, before mounting a cryptanalysis, it is necessary to strip off the data. This reverse-engineering process is often subtle, even when the data formatting is not concealed on purpose. A typical case is interception: some raw data, not necessarily encrypted, is observed out of a noisy channel. To access the information, the whole communication system has first to be disassembled and every constituent reconstructed. A transmission system actually corresponds to a succession of elements (symbol mapping, scrambler, channel encoder, interleaver...), and there exist many possibilities for each of them. In addition to the “preliminary to cryptanalysis” aspect, there are other links between those problems and cryptology. They share some scientific tools (algorithmics, discrete mathematics, probability...), but beyond that, there are some very strong similarities in the techniques.

SPECFUN Project-Team

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.

CAIRN Project-Team

4. Application Domains

4.1. Panorama

keywords: Wireless (Body) Sensor Networks, High-Rate Optical Communications, Wireless Communications, Applied Cryptography, Machine Learning, Deep Learning, Image ans Signal Processing.
Our research is based on realistic applications, in order to both discover the main needs created by these applications and to invent realistic and interesting solutions.

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

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

CAMUS Project-Team

4. Application Domains

4.1. Application Domains

Computational performance being our main objective, our target applications are characterized by intensive computation phases. Such applications are numerous in the domains of scientific computations, optimization, data mining and multimedia.

Applications involving intensive computations are necessarily high energy consumers. However this consumption can be significantly reduced thanks to optimization and parallelization. Although this issue is not our prior objective, we can expect some positive effects for the following reasons:

- Program parallelization tries to distribute the workload equally among the cores. Thus an equivalent performance, or even a better performance, to a sequential higher frequency execution on one single core, can be obtained.
- Memory and memory accesses are high energy consumers. Lowering the memory consumption, lowering the number of memory accesses and maximizing the number of accesses in the low levels of the memory hierarchy (registers, cache memories) have a positive consequence on execution speed, but also on energy consumption.

CASH Project-Team (section vide)

CORSE Project-Team

4. Application Domains

4.1. Transfer

The main industrial sector related to the research activities of CORSE is the one of semi-conductor (programmable architectures spanning from embedded systems to servers). Obviously any computing application which has the objective of exploiting as much as possible the resources (in terms of high-performance but also low energy consumption) of the host architecture is intended to take advantage of advances in compiler and run-time technology. These applications are based over numerical kernels (linear algebra, FFT, convolution...) that can be adapted on a large spectrum of architectures. More specifically, an important activity concerns the optimization of machine learning applications for some high-performance accelerators. Members of CORSE already maintain fruitful and strong collaborations with several companies such as STMICROELECTRONICS, Atos/Bull, Kalray.

PACAP Project-Team

4. Application Domains

4.1. Domains

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

HYCOMES Project-Team (section vide)

Kairos Project-Team

4. Application Domains

4.1. Cyber-Physical and Embedded Systems

We have historical contacts with industrial and academic partners in the domains of avionics and embedded electronics (Airbus, Thales, Safran). We have new collaborations in the fields of satellites (Thales Alenia Space) and connected cars (Renault Software Labs). These provide for use case and new issues in CPS co-modeling and co-design (Digital Twins) further described in New Results section.

4.2. Connected Objects in the Internet Of Things

Due to increasing collaborations with local partners, we have recently considered Smart Contracts (as popularized in Blockchain frameworks), as a way to formally established specification of behavioral system traces, applied to connected objects in a IoT environment. The new ANR project SIM is based on this.

KOPERNIC Team

4. Application Domains

4.1. Avionics

This work is based on a direct collaboration between Airbus and Inria, complementary to collaborative projects like PIA LEOC Capacites and CIFRE thesis. The time critical solutions in this context are based on temporal and spatial isolation of the programs and the understanding of multicore interferences is crucial. Our contributions belong mainly to the solutions space for the objective identified in Section 3.1 .

4.2. Railway

This work is based on a direct collaboration with Clearsy and SNCF, complementary to collaborative projects like PIA BGLE Departs and FUI 21 Waruna. The time critical solutions in this context concern both the proposition of an appropriate scheduler and associated schedulability analyses. Our contributions belong to the solutions space of problems dealt within the objectives identified in Section 3.1 .

4.3. Autonomous cars

This work is based on a direct collaboration with RITS (Inria project team). The time critical solutions in this context concern the interaction between programs executed on multicore processors and messages transmitted through wireless communication channels. Our contributions belong to the solutions space of all three classes of problems dealt within the objectives identified in Section 3.2 .

4.4. Drones

This work is based on the collaborative project FUI/FEDER 22 Ceos. As in the case of autonomous cars, there is an interaction between programs and messages, suggesting that our contributions in this context belong to the solutions space of all three classes of problems dealt within the objectives identified in Section 3.2 .

PARKAS Project-Team

4. Application Domains

4.1. Embedded Control Software

Embedded control software defines the interactions of specialized hardware with the physical world. It normally ticks away unnoticed inside systems like medical devices, trains, aircraft, satellites, and factories. This software is complex and great effort is required to avoid potentially serious errors, especially over many years of maintenance and reuse.

Engineers have long designed such systems using block diagrams and state machines to represent the underlying mathematical models. One of the key insights behind synchronous programming languages is that these models can be executable and serve as the base for simulation, validation, and automatic code generation. This approach is sometimes termed Model-Based Development (MBD). The SCADE language and associated code generator allow the application of MBD in safety-critical applications. They incorporate ideas from LUSTRE, LUCID SYNCHRONE, and other programming languages.

4.2. Hybrid Systems Design and Simulation

Modern embedded systems are increasingly conceived as rich amalgams of software, hardware, networking, and physical processes. The terms Cyberphysical System (CPS) or Internet-of-Things (IoT) are sometimes used as labels for this point of view.

In terms of modeling languages, the main challenges are to specify both discrete and continuous processes in a single *hybrid* language, give meaning to their compositions, simulate their interactions, analyze the behavior of the overall system, and extract code either for target control software or more efficient, possibly online, simulation. Languages like Simulink and Modelica are already used in the design and analysis of embedded systems; it is more important than ever to understand their underlying principles and to propose new constructs and analyses.

SPADES Project-Team

4. Application Domains

4.1. Industrial Applications

Our applications are in the embedded system area, typically: transportation, energy production, robotics, telecommunications, the Internet of things (IoT), systems on chip (SoC). In some areas, safety is critical, and motivates the investment in formal methods and techniques for design. But even in less critical contexts, like telecommunications and multimedia, these techniques can be beneficial in improving the efficiency and the quality of designs, as well as the cost of the programming and the validation processes.

Industrial acceptance of formal techniques, as well as their deployment, goes necessarily through their usability by specialists of the application domain, rather than of the formal techniques themselves. Hence, we are looking to propose domain-specific (but generic) realistic models, validated through experience (*e.g.*, control tasks systems), based on formal techniques with a high degree of automation (*e.g.*, synchronous models), and tailored for concrete functionalities (*e.g.*, code generation).

4.2. Current Industrial Cooperations

Regarding applications and case studies with industrial end-users of our techniques, we cooperate with Orange Labs on software architecture for cloud services.

TEA Project-Team

4. Application Domains

4.1. Automotive and Avionics

From our continuous collaboration with major academic and industrial partners through projects TOPCASED, OPENEMBEDD, SPACIFY, CESAR, OPEES, P and CORAIL, our experience has primarily focused on the aerospace domain. The topics of time and architecture of team TEA extend to both avionics and automotive. Yet, the research focuses on time in team TEA is central in any aspect of, cyber-physical, embedded system design in factory automation, automotive, music synthesis, signal processing, software radio, circuit and system on a chip design; many application domains which, should more collaborators join the team, would definitely be worth investigating.

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

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

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

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

4.2. Factory Automation

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

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

ANTIQUE Project-Team

4. Application Domains

4.1. Verification of safety critical embedded software

The verification of safety critical embedded software is a very important application domain for our group. First, this field requires a high confidence in software, as a bug may cause disastrous events. Thus, it offers an obvious opportunity for a strong impact. Second, such software usually have better specifications and a better design than many other families of software, hence are an easier target for developing new static analysis techniques (which can later be extended for more general, harder to cope with families of programs). This includes avionics, automotive and other transportation systems, medical systems ...

For instance, the verification of avionics systems represent a very high percentage of the cost of an airplane (about 30 % of the overall airplane design cost). The state of the art development processes mainly resort to testing in order to improve the quality of software. Depending on the level of criticality of a software (at the highest levels, any software failure would endanger the flight) a set of software requirements are checked with test suites. This approach is both costly (due to the sheer amount of testing that needs to be performed) and unsound (as errors may go unnoticed, if they do not arise on the test suite).

By contrast, static analysis can ensure higher software quality at a lower cost. Indeed, a static analyzer will catch all bugs of a certain kind. Moreover, a static analysis run typically lasts a few hours, and can be integrated in the development cycle in a seamless manner. For instance, ASTRÉE successfully verified the absence of runtime error in several families of safety critical fly-by-wire avionic software, in at most a day of computation, on standard hardware. Other kinds of synchronous embedded software have also been analyzed with good results.

In the future, we plan to greatly extend this work so as to verify *other families of embedded software* (such as communication, navigation and monitoring software) and *other families of properties* (such as security and liveness properties).

Embedded software in charge of communication, navigation, and monitoring typically relies on a *parallel* structure, where several threads are executed concurrently, and manage different features (input, output, user interface, internal computation, logging ...). This structure is also often found in automotive software. An even more complex case is that of *distributed* systems, where several separate computers are run in parallel and take care of several sub-tasks of a same feature, such as braking. Such a logical structure is not only more complex than the synchronous one, but it also introduces new risks and new families of errors (deadlocks, data-races...). Moreover, such less well designed, and more complex embedded software often utilizes more complex data-structures than synchronous programs (which typically only use arrays to store previous states) and may use dynamic memory allocation, or build dynamic structures inside static memory regions, which are actually even harder to verify than conventional dynamically allocated data structures. Complex data-structures also introduce new kinds of risks (the failure to maintain structural invariants may lead to runtime errors, non termination, or other software failures). To verify such programs, we will design additional abstract domains, and develop new static analysis techniques, in order to support the analysis of more complex programming language features such as parallel and concurrent programming with threads and manipulations of complex data structures. Due to their size and complexity, the verification of such families of embedded software is a major challenge for the research community.

Furthermore, embedded systems also give rise to novel security concerns. It is in particular the case for some aircraft-embedded computer systems, which communicate with the ground through untrusted communication media. Besides, the increasing demand for new capabilities, such as enhanced on-board connectivity, e.g. using mobile devices, together with the need for cost reduction, leads to more integrated and interconnected systems. For instance, modern aircrafts embed a large number of computer systems, from safety-critical cockpit avionics to passenger entertainment. Some systems meet both safety and security requirements.

Despite thorough segregation of subsystems and networks, some shared communication resources raise the concern of possible intrusions. Because of the size of such systems, and considering that they are evolving entities, the only economically viable alternative is to perform automatic analyses. Such analyses of security and confidentiality properties have never been achieved on large-scale systems where security properties interact with other software properties, and even the mapping between high-level models of the systems and the large software base implementing them has never been done and represents a great challenge. Our goal is to prove empirically that the security of such large scale systems can be proved formally, thanks to the design of dedicated abstract interpreters.

The long term goal is to make static analysis more widely applicable to the verification of industrial software.

4.2. Static analysis of software components and libraries

An important goal of our work is to make static analysis techniques easier to apply to wider families of software. Then, in the longer term, we hope to be able to verify less critical, yet very commonly used pieces of software. Those are typically harder to analyze than critical software, as their development process tends to be less rigorous. In particular, we will target operating systems components and libraries. As of today, the verification of such programs is considered a major challenge to the static analysis community.

As an example, most programming languages offer Application Programming Interfaces (API) providing ready-to-use abstract data structures (e.g., sets, maps, stacks, queues, etc.). These APIs, are known under the name of containers or collections, and provide off-the-shelf libraries of high level operations, such as insertion, deletion and membership checks. These container libraries give software developers a way of abstracting from low-level implementation details related to memory management, such as dynamic allocation, deletion and pointer handling or concurrency aspects, such as thread synchronization. Libraries implementing data structures are important building bricks of a huge number of applications, therefore their verification is paramount. We are interested in developing static analysis techniques that will prove automatically the correctness of large audience libraries such as Glib and Threading Building Blocks.

4.3. Models of mechanistic interactions between proteins

Computer Science takes a more and more important role in the design and the understanding of biological systems such as signaling pathways, self assembly systems, DNA repair mechanisms. Biology has gathered large data-bases of facts about mechanistic interactions between proteins, but struggles to draw an overall picture of how these systems work as a whole. High level languages designed in Computer Science allow one to collect these interactions in integrative models, and provide formal definitions (i.e., semantics) for the behavior of these models. This way, modelers can encode their knowledge, following a bottom-up discipline, without simplifying *a priori* the models at the risk of damaging the key properties of the system. Yet, the systems that are obtained this way suffer from combinatorial explosion (in particular, in the number of different kinds of molecular components, which can arise at run-time), which prevents from a naive computation of their behavior.

We develop various analyses based on abstract interpretation, and tailored to different phases of the modeling process. We propose automatic static analyses in order to detect inconsistencies in the early phases of the modeling process. These analyses are similar to the analysis of classical safety properties of programs. They involve both forward and backward reachability analyses as well as causality analyses, and can be tuned at different levels of abstraction. We also develop automatic static analyses in order to identify key elements in the dynamics of these models. The results of these analyses are sent to another tool, which is used to automatically simplify models. The correctness of this simplification process is proved by the means of abstract interpretation: this ensures formally that the simplification preserves the quantitative properties that have been specified beforehand by the modeler. The whole pipeline is parameterized by a large choice of abstract domains which exploits different features of the high level description of models.

4.4. Consensus

Fault-tolerant distributed systems provide a dependable service on top of unreliable computers and networks. Famous examples are geo-replicated data-bases, distributed file systems, or blockchains. Fault-tolerant protocols replicate the system and ensure that all (unreliable) replicas are perceived from the outside as one single reliable machine. To give the illusion of a single reliable machine "consensus" protocols force replicas to agree on the "current state" before making this state visible to an outside observer. We are interested in (semi-)automatically proving the total correctness of consensus algorithms in the benign case (messages are lost or processes crash) or the Byzantine case (processes may lie about their current state). In order to do this, we first define new reduction theorems to simplify the behaviors of the system and, second, we introduce new static analysis methods to prove the total correctness of adequately simplified systems. We focus on static analysis based Satisfiability Modulo Theories (SMT) solvers which offers a good compromise between automation and expressiveness. Among our benchmarks are Paxos, PBFT (Practical Byzantine Fault-Tolerance), and blockchain algorithms (Red-Belly, Tendermint, Algorand). These are highly challenging benchmarks, with a lot of non-determinism coming from the interleaving semantics and from the adversarial environment in which correct processes execute, environment that can drop messages, corrupt them, etc. Moreover, these systems were originally designed for a few servers but today are deployed on networks with thousands of nodes. The "optimizations" for scalability can no longer be overlooked and must be considered as integral part of the algorithms, potentially leading to specifications weaker than the so much desired consensus.

4.5. Models of growth

In systems and synthetic biology (engineered systems) one would like study the environment of a given cellular process (such as signaling pathways mentioned earlier) and the ways in which that process interacts with different resources provided by the host. To do this, we have built coarse-grained models of cellular physiology which summarize fundamental processes (transcription, translation, transport, metabolism). such models describe global growth in mechanistic way and allow one to plug the model of one's process of interest into a simplified and yet realistic and reactive model of the process interaction with its immediate environment. A first ODE-based deterministic version of this model [26] explaining the famous bacterial growth laws and how the allocation of resources to different genomic sectors depends on the growth conditions- was published in 2015 and has already received nearly 150 citations. The model also allows one to bridge between population genetic models which describe cells in terms of abstract features and fitness and intra-cellular models. For instance, we find that fastest growing strategies are not evolutionary stable in competitive experiments. We also find that vastly different energy storage strategies exist [24]. In a recent article [25] in *Nature Communications* we build a stochastic version of the above model. We predict the empirical size and doubling time distributions as a function of growth conditions. To be able to fit the parameters of the model to available single-cell data (note that the fitting constraints are far tighter than in the deterministic case), we introduce new techniques for the approximation of reaction-division systems which generalize continuous approximations of Langevin type commonly used for pure reaction systems. We also use cross-correlations to visualize causality and modes in noise propagation in the model (in a way reminiscent to abstract computational traces mentioned earlier). In other work, we show how to connect our new class of models to more traditional ones stemming from "flux balance analysis" by introducing an allocation vector which allows one to assign a formal growth rate to a class of reaction systems [20].

4.6. Static analysis of data science software

Nowadays, thanks to advances in machine learning and the availability of vast amounts of data, computer software plays an increasingly important role in assisting or even autonomously performing tasks in our daily lives. As data science software becomes more and more widespread, we become increasingly vulnerable to programming errors. In particular, programming errors that do not cause failures can have serious consequences since code that produces an erroneous but plausible result gives no indication that something went wrong. This issue becomes particularly worrying knowing that machine learning software, thanks to its ability

to efficiently approximate or simulate more complex systems, is slowly creeping into mission critical scenarios. However, programming errors are not the only concern. Another important issue is the vulnerability of machine learning models to adversarial examples, that is, small input perturbations that cause the model to misbehave in unpredictable ways. More generally, a critical issue is the notorious difficulty to interpret and explain machine learning software. Finally, as we are witnessing widespread adoption of software with far-reaching societal impact — i.e., to automate decision-making in fields such as social welfare, criminal justice, and even health care — a number of recent cases have evidenced the importance of ensuring software fairness as well as data privacy. Going forward, data science software will be subject to more and more legal regulations (e.g., the European General Data Protection Regulation adopted in 2016) as well as administrative audits. It is thus paramount to develop method and tools that can keep up with these developments and enhance our understanding of data science software and ensure it behaves correctly and reliably. In particular, we are interesting in developing new static analyses specifically tailored to the idiosyncrasies of data science software. This makes it a new and exciting area for static analysis, offering a wide variety of challenging problems with huge potential impact on various interdisciplinary application domains [13].

CAMBIUM Project-Team

4. Application Domains

4.1. Formal methods

We develop techniques and tools for the formal verification of critical software:

- program logics based on CFML and Iris for the deductive verification of software, including concurrency and algorithmic complexity aspects;
- verified development tools such as the CompCert verified C compiler, which extends properties established by formal verification at the source level all the way to the final executable code.

Some of these techniques have already been used in the nuclear industry (MTU Friedrichshafen uses CompCert to develop emergency diesel generators) and are under evaluation in the aerospace industry.

4.2. High-assurance software

Software that is not critical enough to undergo formal verification can still benefit greatly, in terms of reliability and security, from a functional, statically-typed programming language. The OCaml type system offers several advanced tools (generalized algebraic data types, abstract types, extensible variant and object types) to express many data structure invariants and safety properties and have them automatically enforced by the type-checker. This makes OCaml a popular language to develop high-assurance software, in particular in the financial industry. OCaml is the implementation language for the Tezos blockchain and cryptocurrency. It is also used for automated trading at Jane Street and for modeling and pricing of financial contracts at Bloomberg, Lexifi and Simcorp. OCaml is also widely used to implement code verification and generation tools at Facebook, Microsoft, CEA, Esterel Technologies, and many academic research groups, at Inria and elsewhere.

4.3. Design and test of microprocessors

The **diy** tool suite and the underlying methodology is in use at ARM Ltd to design and test the memory model of ARM architectures. In particular, the internal reference memory model of the ARMv8 (or AArch64) architecture has been written “in house” in Cat, our domain-specific language for specifying and simulating memory models. Moreover, our test generators and runtime infrastructure are used routinely at ARM to test various implementations of their architectures.

4.4. Teaching programming

Our work on the OCaml language family has an impact on the teaching of programming. OCaml is one of the programming languages selected by the French Ministry of Education for teaching Computer Science in classes préparatoires scientifiques. OCaml is also widely used for teaching advanced programming in engineering schools, colleges and universities in France, the USA, and Japan. The MOOC “Introduction to Functional Programming in OCaml”, developed at University Paris Diderot, is available on the France Université Numérique platform and comes with an extensive platform for self-training and automatic grading of exercises, developed in OCaml itself.

CELTIQUE Project-Team (section vide)

CONVECS Project-Team

4. Application Domains

4.1. Application Domains

The theoretical framework we use (automata, process algebras, bisimulations, temporal logics, etc.) and the software tools we develop are general enough to fit the needs of many application domains. They are applicable to virtually any system or protocol that consists of distributed agents communicating by asynchronous messages. The list of recent case studies performed with the CADP toolbox (see in particular § 7.5) illustrates the diversity of applications:

- *Bioinformatics:* genetic regulatory networks, nutritional stress response, metabolic pathways,
- *Component-based systems:* Web services, peer-to-peer networks,
- *Cloud computing:* self-deployment protocols, dynamic reconfiguration protocols,
- *Fog and IoT:* stateful IoT applications in the fog,
- *Databases:* transaction protocols, distributed knowledge bases, stock management,
- *Distributed systems:* virtual shared memory, dynamic reconfiguration algorithms, fault tolerance algorithms, cloud computing,
- *Embedded systems:* air traffic control, avionic systems, train supervision systems, medical devices,
- *Hardware architectures:* multiprocessor architectures, systems on chip, cache coherency protocols, hardware/software codesign,
- *Human-machine interaction:* graphical interfaces, biomedical data visualization, plasticity,
- *Security protocols:* authentication, electronic transactions, cryptographic key distribution,
- *Telecommunications:* high-speed networks, network management, mobile telephony, feature interaction detection.

DEDUCTEAM Project-Team

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.

GALLINETTE Project-Team (section vide)

MEXICO Project-Team

4. Application Domains

4.1. Telecommunications

Participants: 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 Regulation Networks

Participants: Thomas Chatain, Matthias Fuegger, Stefan Haar, Serge Haddad, Juraj Kolcak, Hugues Mandon, 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 Manish Kushwaha (INRA), and Thomas Nowak (LRI) on biological infection models for E. coli colonies and M13 phages.

4.3. Metabolic Networks

Participant: Philippe Dague.

Analysis of metabolic networks in presence of biological (thermodynamical, kinetic, gene regulatory) constraints has been studied achieving a complete mathematical characterization of the solutions space at steady state (generalization of the elementary flux modes) and investigating related computing methods.

4.4. Transportation Systems

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

- **Autonomous Vehicles.** 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.
- **Multimodal Transport Networks.** We are interested in predicting and harnessing the propagation of perturbations across different transportation modes.

MOCQUA Team

4. Application Domains

4.1. Quantum Computing

Quantum Computing is currently the most promising technology to extend Moore's law, whose end is expected with the engraving at 7 nm, in less than 5 years. Thanks to the exponential computational power it will bring, it will represent a decisive competitive advantage for those who will control it.

Quantum Computing is also a major security issue, since it allows us to break today's asymmetric cryptography. Hence, mastering quantum computing is also of the highest importance for national security concerns. Recent scientific and technical advances suggest that the construction of the first quantum computers will be possible in the coming years.

As a result, the major US players in the IT industry have embarked on a dramatic race, mobilizing huge resources: IBM, Microsoft, Google and Intel have each invested between 20 and 50 million euros, and are devoting significant budgets to attract and hire the best scientists on the planet. Some states have launched ambitious national programs, including Great Britain, the Netherlands, Canada, China, Australia, Singapore, and very recently Europe, with the upcoming 10-year FET Flagship program in Quantum Engineering.

While a large part of these resources are going towards R-&-D in quantum hardware, there is still an important need and real opportunities for leadership in the field of quantum software.

The Mocqua team contributes to the computer science approach to quantum computing, aka the quantum software approach. We aim at a better understanding of the power and limitations of the quantum computer, and therefore of its impact on society. We also contribute to ease the development of the quantum computer by filling the gap between the theoretical results on quantum algorithms and complexity and the recent progresses in quantum hardware.

4.2. Higher-Order Computing

The idea of considering functions as first-class citizens and allowing programs to take functions as inputs has emerged since the very beginning of theoretical computer science through Church's λ-calculus and is nowadays at the core of functional programming, a paradigm that is used in modern software and by digital companies (Google, Facebook, ...). In the meantime higher-order computing has been explored in many ways in the fields of logic and semantics of programming languages.

One of the central problems is to design programming languages that capture most of, if not all, the possible ways of computing with functions as inputs. There is no Church thesis in higher-order computing and many ways of taking a function as input can be considered: allowing parallel or only sequential computations, querying the input as a black-box or via an interactive dialog, and so on.

The Kleene-Kreisel computable functionals are arguably the broadest class of higher-order continuous functionals that could be computed by a machine. However their complexity is such that no current programming language can capture all of them. Better understanding this class of functions is therefore fundamental in order to identify the features that a programming language should implement to make the full power of higher-order computation expressible in such a language.

4.3. Simulation of Dynamical Systems by Cellular Automata

We aim at developing various tools to simulate and analyse the dynamics of spatially-extended discrete dynamical systems such as cellular automata. The emphasis of our approach is on the evaluation of the robustness of the models under study, that is, their capacity to resist various perturbations.

In the framework of pure computational questions, various examples of such systems have already been proposed for solving complex problems with a simple bio-inspired approach (e.g. the decentralized gathering problem [40]). We are now working on their transposition to various real-world situations. For example when one needs to understand the behaviour of large-scale networks of connected components such as wireless sensor networks. In this direction of research, a first work has been presented on how to achieve a decentralized diagnosis of networks made of simple interacting components and the results are rather encouraging [5]. Nevertheless, there are various points that remain to be studied in order to complete this model for its integration in a real network.

We have also tackled the question of the evaluation of the robustness of a swarming model proposed by A. Deutsch to mimic the self-organization process observed in various natural systems (birds, fishes, bacteria, etc.) [2]. We now wish to develop our simulation tools to apply them to various biological phenomena where a great number of agents are implied.

We are also currently extending the range of applications of these techniques to the field of economy. We have started a collaboration with Massimo Amato, a professor in economy at the Bocconi University in Milan. Our aim is to examine how to propose a decentralized view of a business-to-business market and propose agent-oriented and totally decentralized models of such markets. Various banks and large businesses have already expressed their interest in such modelling approaches.

PARSIFAL Project-Team

4. Application Domains

4.1. Automated Theorem Proving

The Parsifal team studies the structure of mathematical proofs, in ways that often makes them more amenable to automated theorem proving – automatically searching the space of proof candidates for a statement to find an actual proof – or a counter-example.

(Due to fundamental computability limits, fully-automatic proving is only possible for simple statements, but this field has been making a lot of progress in recent years, and is in particular interested with the idea of generating verifiable evidence for the proofs that are found, which fits squarely within the expertise of Parsial.)

4.2. Proof-assistants

The team work on the structure of proofs also suggests ways that they could be presented to a user, edited and maintained, in particular in “proof assistants”, automated tool to assist the writing of mathematical proofs with automatic checking of their correctness.

4.3. Programming language design

Our work also gives insight on the structure and properties of programming languages. We can improve the design or implementation of programming languages, help programmers or language implementors reason about the correctness of the programs in a given language, or reason about the cost of execution of a program.

PI.R2 Project-Team (section vide)

STAMP Project-Team

4. Application Domains

4.1. Mathematical Components

The Mathematical Components is the main by-product of an effort started almost two decades ago to provide a formally verified proof for a major theorem in group theory. Because this major theorem had a proof published in books of several hundreds of pages, with elements coming from character theory, other coming from algebra, and some coming from real analysis, it was an exercice in building a large library, with results in many domains, and in establishing clear guidelines for further increase and data search.

This library has proved to be a useful repository of mathematical facts for a wide area of applications, so that it has a growing community of users in many contries (Denmark, France, Germany, Japan, Singapore, Spain, Sweden, UK, USA, at the time of writing these lines in 2019) and for a wide variety of topics (transcendental number theory, elliptic curve cryptography, articulated robot kinematics, recently block chain foundations).

Interesting questions on this library range around the importance of decidability and proof irrelevance, the way to structure knowledge to automatically inherit theorems from one topic to another, the way to generate infrastructure to make this automation efficient and predictable. In particular, we want to concentrate on adding a new mathematical topic to this library: real analysis and then complex analysis (Mathematical Components Analysis).

On the front of automation, we are convinced that a higher level language is required to describe similarities between theories, to generate theorems that are immediate consequences of structures, etc, and for this reason, we invest in the development of a new language on top of the proof assistant (ELPI).

4.2. Proofs in cryptography

When we work on cryptography, we are interested in the formal verification of proofs showing that some cryptographic primitives provide good guarantees against unwanted access to information. Over the years we have developed a technique for this kind of reasoning that relies on a programing logic (close to Hoare logic) with probabilistic apsects and the capability to establish relations between several implementations of a problem. The resulting programming logic is called *probabilistic relational Hoare logic*. In more recent work, we have also started to study questions of *side-channel* attacks, where we wish to guarantee that opponents cannot gain access to protected knowledge, even if they observe specific features of execution, like execution time (to which the answer lies in *constant-time* execution) or partial access to memory bits (to which the answer lies in *masking*).

For this domain of application, we choose to work with a specific proof tool (EasyCrypt), which combines powerful first-order reasoning and uses of automatic tools, with a specific support for probabilistic relational Hoare Logic. The development of this EasyCrypt proof tool is one of the objectives of our team.

When it comes to formal proofs of resistance to side-channel attack, we contend that it is necessary to verify formally that the compiler used in the production of actually running code respects the resistance properties that were established in formally verified proofs. One of our objectives is to describe such a compiler (Jasmin) and show its strength on a variety of applications.

4.3. Proofs for robotics

Robots are man-made artifacts where numerous design decisions can be argued based on logical or mathematical principles. For this reason, we wish to use this domain of application as a focus for our investigations. The questions for which we are close to providing answers involve precision issues in numeric computation, obstacle avoidance and motion planning (including questions of graph theory), articulated limb cinematics and dynamics, and balance and active control.

From the mathematical perspective, these topics require that we improve our library to cover real algebraic geometry, computational geometry, real analysis, graph theory, and refinement relations between abstract algorithms and executable programs.

In the long run, we hope to exhibit robots where pieces of software and part of the design has been subject to formal verification.

SUMO Project-Team

4. Application Domains

4.1. Smart transportation systems

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

4.2. Management of telecommunication networks and of data centers

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

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

4.3. Collaborative workflows

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

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

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

TOCCATA Project-Team

4. Application Domains

4.1. Safety-Critical Software

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 or impact mainly belong to the domain of transportation: aerospace, aviation, railway, automotive.

Transfer to aeronautics: Airbus France Development of the control software of Airbus planes historically includes advanced usage of formal methods. A first aspect is the usage of the CompCert verified compiler for compiling C source code. Our work in cooperation with Gallium team for the safe compilation of floating-point arithmetic operations [2] is directly in application in this context. A second aspect is the usage of the Frama-C environment for static analysis to verify the C source code. In this context, both our tools Why3 and Alt-Ergo are indirectly used to verify C code.

Transfer to the community of Atelier B In the former ANR project BWare, we investigated 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 https://www.atelierb.eu/en/. The transfer effort continues nowadays through the FUI project LCHIP.

ProofInUse joint lab: transfer to the community of Ada development Through the creation of the ProofInUse joint lab (https://www.adacore.com/proofinuse) in 2014, with AdaCore company (https://www.adacore.com/), we have a growing impact on the community of industrial development of safety-critical applications written in Ada. See the web page https://www.adacore.com/industries for a an overview of AdaCore's customer projects, in particular those involving the use of the SPARK Pro tool set. This impact involves both the use of Why3 for generating VCs on Ada source codes, and the use of Alt-Ergo for performing proofs of those VCs.

The impact of ProofInUse can also be measured in term of job creation: the first two ProofInUse engineers, D. Hauzar and C. Fumex, employed initially on Inria temporary positions, have now been hired on permanent positions in AdaCore company. It is also interesting to notice that this effort allowed AdaCore company to get new customers, in particular the domains of application of deductive formal verification went beyond the historical domain of aerospace: application in automotive (https://www.adacore.com/customers/toyota-itc-japan) cyber-security (https://www.adacore.com/customers/multi-level-security-workstation), health (artificial heart, https://www.adacore.com/customers/total-artificial-heart).

Extension of ProofInUse joint lab The current plans for continuation of the ProofInUse joint lab (https://why3.gitlabpages.inria.fr/proofinuse/) include extension at a larger perimeter than Ada applications. We started to collaborate with the TrustInSoft company (https://trust-in-soft.com/) for the verification of C and C++ codes, including the use of Why3 to design verified and reusable C libraries (ongoing CIFRE PhD thesis). We also started to collaborate with Mitsubishi Electric in Rennes (http://www.mitsubishielectric-rce.eu/xindex.php) for a specific usage of Why3 for verifying embedded devices (logic controllers).

Generally speaking, we believe that our increasing industrial impact is a representative success for our general goal of spreading deductive verification methods to a larger audience, and we are firmly engaged into continuing such kind of actions in the future.

VERIDIS Project-Team

4. Application Domains

4.1. Application Domains

Distributed algorithms and protocols are found at all levels of computing infrastructure, from many-core processors and systems-on-chip to wide-area networks. We are particularly interested in the verification of algorithms that are developed for supporting novel computing paradigms, including ad-hoc networks that underly mobile and low-power computing or overlay networks, peer-to-peer networks that provide services for telecommunication, or cloud computing services. Computing infrastructure must be highly available and is ideally invisible to the end user, therefore correctness is crucial. One should note that standard problems of distributed computing such as consensus, group membership or leader election have to be reformulated for the dynamic context of these modern systems. We are not ourselves experts in the design of distributed algorithms, but we work together with domain experts on designing formal models of these protocols, and on verifying their properties. These cooperations help us focus on concrete algorithms and ensure that our work is relevant to the distributed algorithm community.

Our work on symbolic procedures for solving polynomial constraints finds applications beyond verification. In particular, we have been working in interdisciplinary projects with researchers from mathematics, computer science, system biology, and system medicine on the analysis of molecular interaction networks in order to infer the principal qualitative properties of models. Our techniques complement numerical analysis techniques and are validated against collections of models from computational biology.

CIDRE Project-Team (section vide)

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.

DATASPHERE Team

4. Application Domains

4.1. Governance

- City governance, local democracy and interaction with citizens.
- Local governance versus global norms and control.
- Strategy beyond public open data.
- Smart city governance.

4.2. CyberSecurity

- Cyber-strategy, defense and security in an evolving world shaped by the digital in particular China/Russia/US cyber-strategy.
- Data strategy for the digital economy, cross border intermediation, platform strategie.
- Strategy of Artificial Intelligence, transparency/acceptability/explainability of AI.
- Cartography of the cyberspace.
- Network, BGP security.

4.3. Anthropocene

- Adaptation to the conditions of the anthropocene, digital control of resources and homeostasis.
- Geopolitics of the environmental challenges, adaptation and mitigation.
- Contemporaneity of the digital revolution and global warming.

PESTO Project-Team

4. Application Domains

4.1. Cryptographic protocols

Security protocols, such as TLS, Kerberos, ssh or AKA (mobile communication), are the main tool for securing our communications. The aim of our work is to improve their security guarantees. For this, we propose models that are expressive enough to formally represent protocol executions in the presence of an adversary, formal definitions of the security properties to be satisfied by these protocols, and automated tools able to analyse them and possibly exhibit design flaws.

4.2. Automated reasoning

Many techniques for symbolic verification of security are rooted in automated reasoning. A typical example is equational reasoning used to model the algebraic properties of a cryptographic primitive. Our work therefore aims to improve and adapt existing techniques or propose new ones when needed for reasoning about security.

4.3. Electronic voting

Electronic elections have in the last years been used in several countries for politically binding elections. The use in professional elections is even more widespread. The aim of our work is to increase our understanding of the security properties needed for secure elections, propose techniques for analysing e-voting protocols, design of state-of-the-art voting protocols, but also to highlight the limitations of e-voting solutions.

4.4. Privacy in social networks

The treatment of information released by users on social networks can violate a user's privacy. The goal of our work is to allow users to control the information released while guaranteeing their privacy.

PRIVATICS Project-Team

3. Application Domains

3.1. Domain 1: Privacy in smart environments

Privacy in smart environments. One illustrative example is our latest work on privacy-preserving smart-metering [2]. Several countries throughout the world are planning to deploy smart meters in house-holds in the very near future. Traditional electrical meters only measure total consumption on a given period of time (i.e., one month or one year). As such, they do not provide accurate information of when the energy was consumed. Smart meters, instead, monitor and report consumption in intervals of few minutes. They allow the utility provider to monitor, almost in real-time, consumption and possibly adjust generation and prices according to the demand. Billing customers by how much is consumed and at what time of day will probably change consumption habits to help matching energy consumption with production. In the longer term, with the advent of smart appliances, it is expected that the smart grid will remotely control selected appliances to reduce demand. Although smart metering might help improving energy management, it creates many new privacy problems. Smart-meters provide very accurate consumption data to electricity providers. As the interval of data collected by smart meters decreases, the ability to disaggregate low-resolution data increases. Analysing high-resolution consumption data, Non-intrusive Appliance Load Monitoring (NALM) can be used to identify a remarkable number of electric appliances (e.g., water heaters, well pumps, furnace blowers, refrigerators, and air conditioners) employing exhaustive appliance signature libraries. We developed DREAM, DiffeRentially privatE smArt Metering, a scheme that is private under the differential privacy model and therefore provides strong and provable guarantees. With our scheme, an (electricity) supplier can periodically collect data from smart-meters and derive aggregated statistics while learning only limited information about the activities of individual households. For example, a supplier cannot tell from a user's trace when he watched TV or turned on heating.

3.2. Domain 2: Big Data and Privacy

We believe that another important problem will be related to privacy issues in big data. Public datasets are used in a variety of applications spanning from genome and web usage analysis to location-based and recommendation systems. Publishing such datasets is important since they can help us analyzing and understanding interesting patterns. For example, mobility trajectories have become widely collected in recent years and have opened the possibility to improve our understanding of large-scale social networks by investigating how people exchange information, interact, and develop social interactions. With billion of handsets in use worldwide, the quantity of mobility data is gigantic. When aggregated, they can help understand complex processes, such as the spread of viruses, and build better transportation systems. While the benefits provided by these datasets are indisputable, they unfortunately pose a considerable threat to individual privacy. In fact, mobility trajectories might be used by a malicious attacker to discover potential sensitive information about a user, such as his habits, religion or relationships. Because privacy is so important to people, companies and researchers are reluctant to publish datasets by fear of being held responsible for potential privacy breaches. As a result, only very few of them are actually released and available. This limits our ability to analyze such data to derive information that could benefit the general public. It is now an urgent need to develop Privacy-Preserving Data Analytics (PPDA) systems that collect and transform raw data into a version that is immunized against privacy attacks but that still preserves useful information for data analysis. This is one of the objectives of Privatics. There exists two classes of PPDA according to whether the entity that is collecting and anonymizing the data is trusted or not. In the trusted model, that we refer to as Privacy-Preserving Data Publishing (PPDP), individuals trust the publisher to which they disclose their data. In the untrusted model, that we refer to as Privacy-Preserving Data Collection (PPDC), individuals do not trust the data publisher. They may add some noise to their data to protect sensitive information from the data publisher.

Privacy-Preserving Data Publishing: In the trusted model, individuals trust the data publisher and disclose all their data to it. For example, in a medical scenario, patients give their true information to hospitals to receive proper treatment. It is then the responsibility of the data publisher to protect privacy of the individuals' personal data. To prevent potential data leakage, datasets must be sanitized before possible release. Several proposals have been recently proposed to release private data under the Differential Privacy model [25, 56, 26, 57, 50]. However most of these schemes release a "snapshot" of the datasets at a given period of time. This release often consists of histograms. They can, for example, show the distributions of some pathologies (such as cancer, flu, HIV, hepatitis, etc.) in a given population. For many analytics applications, "snapshots" of data are not enough, and sequential data are required. Furthermore, current work focusses on rather simple data structures, such as numerical data. Release of more complex data, such as graphs, are often also very useful. For example, recommendation systems need the sequences of visited websites or bought items. They also need to analyse people connection graphs to identify the best products to recommend. Network trace analytics also rely on sequences of events to detect anomalies or intrusions. Similarly, traffic analytics applications typically need sequences of visited places of each user. In fact, it is often essential for these applications to know that user A moved from position 1 to position 2, or at least to learn the probability of a move from position 1 to position 2. Histograms would typically represent the number of users in position 1 and position 2, but would not provide the number of users that moved from position 1 to position 2. Due to the inherent sequentiality and high-dimensionality of sequential data, one major challenge of applying current data sanitization solutions on sequential data comes from the uniqueness of sequences (e.g., very few sequences are identical). This fact makes existing techniques result in poor utility. Schemes to privately release data with complex data structures, such as sequential, relational and graph data, are required. This is one the goals of Privatics. In our current work, we address this challenge by employing a variable-length n-gram model, which extracts the essential information of a sequential database in terms of a set of variable-length n - grams [15]. We then intend to extend this approach to more complex data structures.

Privacy-Preserving Data Collection: In the untrusted model, individuals do not trust their data publisher. For example, websites commonly use third party web analytics services, such as Google Analytics to obtain aggregate traffic statistics such as most visited pages, visitors' countries, etc. Similarly, other applications, such as smart metering or targeted advertising applications, are also tracking users in order to derive aggregated information about a particular class of users. Unfortunately, to obtain this aggregate information, services need to track users, resulting in a violation of user privacy. One of our goals is to develop Privacy-Preserving Data Collection solutions. We propose to study whether it is possible to provide efficient collection/aggregation solutions without tracking users, i.e. without getting or learning individual contributions.

PROSECCO Project-Team

4. Application Domains

4.1. Cryptographic Protocol Libraries

Cryptographic protocols such as TLS, SSH, IPSec, and Kerberos are the trusted base on which the security of modern distributed systems is built. Our work enables the analysis and verification of such protocols, both in their design and implementation. Hence, for example, we build and verify models and reference implementations for well-known protocols such as TLS and SSH, as well as analyze their popular implementations such as OpenSSL.

4.2. Hardware-based security APIs

Cryptographic devices such as Hardware Security Modules (HSMs) and smartcards are used to protect long-terms secrets in tamper-proof hardware, so that even attackers who gain physical access to the device cannot obtain its secrets. These devices are used in a variety of scenarios ranging from bank servers to transportation cards (e.g. Navigo). Our work investigates the security of commercial cryptographic hardware and evaluates the APIs they seek to implement.

4.3. Web application security

Web applications use a variety of cryptographic techniques to securely store and exchange sensitive data for their users. For example, a website may serve pages over HTTPS, authenticate users with a single sign-on protocol such as OAuth, encrypt user files on the server-side using XML encryption, and deploy client-side cryptographic mechanisms using a JavaScript cryptographic library. The security of these applications depends on the public key infrastructure (X.509 certificates), web browsers' implementation of HTTPS and the same origin policy (SOP), the semantics of JavaScript, HTML5, and their various associated security standards, as well as the correctness of the specific web application code of interest. We build analysis tools to find bugs in all these artifacts and verification tools that can analyze commercial web applications and evaluate their security against sophisticated web-based attacks.

TAMIS Project-Team (section vide)