RESEARCH CENTER

FIELD
Algorithmics, Programming, Software and Architecture

Activity Report 2018

Section Application Domains

Edition: 2019-03-07

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 [17] 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 [18], 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 [28] is one of only two academic works cited by NIST in the initial version (2011) of the report [31].

We also develop more specialized software. Our flagship software package is Cado-NFS, 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

4. Application Domains

4.1. Main application domains

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

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

GAIA Team

4. Application Domains

4.1. Adaptive & parametric robust control – collaboration with Safran Electronics & Defense

We have developed a collaboration with *Safran Electronics & Defense* (Massy Palaiseau) and Rouillier (OURAGAN, Inria Paris) on a *parametric robust control theory* based on computer algebra methods (symbolic-numeric methods), as well as its applications to the robust stabilization of certain mechanical systems (e.g. gyrostabilized systems, two mass-spring-damper system, stabilized mirrors).

For low-dimensional systems of ODEs, this approach aims to determine closed-form solutions for robust controllers and for the robustness margins in terms of the model parameters (e.g. mass, length, inertia, mode) [12], [98], [100]. The main applications of these results are twofold: the feasibility of an industrial project can be simplified by speeding up the computation of robust controllers and robust margins for systems with rapidly changing architecture parameters, and avoiding usual time-consuming optimization techniques. Secondly, adaptive and embeddable schemes for robust controllers can be proposed and tested while coupling our approach with real-time parameter estimation methods such as the ones developed in the GAIA team. For more details, see [12].

Preliminary works in the direction have opened a great variety of questions such as the explicit search for positive definite solutions of algebraic or differential Riccati equations (i.e. polynomial or differential systems) with model parameters, the reduction of these equations, and of the parameters based on symmetries, the development, of efficient tools for plotting high degree curves and surfaces showing the robustness margins in terms of the model parameters (collaboration with Moroz (GAMBLE, Inria Nancy)), the use of a certified numeric Newton-Puiseux algorithm for the design of robust controllers, etc. [12], [98], [100]. These results require the use of a large spectrum of computer algebra methods such as linear algebra with parameters, polynomial systems with parameters, ordinary differential systems with parameters, symmetries and reduction, rational parametrizations, discriminant varieties, semi-algebraic sets, critical point methods, real root isolation methods, etc. We shall further develop the parametric robust control in collaboration with *Safran Electronics & Defense*.

In connection with the above results, parameter estimation methods will be studied to develop *adaptive robust controllers* for gyrostabilized systems. Indeed, combining explicit characterizations of robust controllers in terms of the model parameters with time-to-time estimations of these model parameters (which can change with the system production, the heat, the wear, etc.), the robust controllers can then be automatically tuned to conserve their robustness performances [12], [99].

Finally, as explained in [11], [99], constant and distributed delays naturally appear in *Safran E & D* systems (e.g. gyrostabilized systems using visual trackers, stabilized mirror models). Extensions of the above problems and results will be studied for differential time-delay systems based on robust control techniques for infinite-dimensional systems (see, e.g., [54] and the references therein) and its algebraic extension to include model parameters.

4.2. Self calibration problem & Gear fault diagnosis – collaboration with Safran Tech

4.2.1. Self calibration problem

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*, a well-known problem in signal processing, 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 two 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.

Based on computer algebra methods for polynomial systems, we have recently proposed a new approach for the MDU problem which yields closed-form solutions and an efficient algorithm for the estimation of the positions [56] only based on linear algebra techniques. This first result, obtained in collaboration with Dagher (Research Engineer, Inria Chile) and Zheng (DEFROST, Inria Lille - Nord Europe), yields a recent *patent* [55]. Real tests are now carried out. Our first results will be further developed, improved, tested, and demonstrated.

The MDU problem is just one instance of localization problems: more problems can be addressed for which a computer algebra expertise can brought new interesting results, especially in finding closed-form solutions, yielding new estimation techniques which avoid the use of optimization algorithms as commonly done in the signal processing literature. The main differences between these localization problems can essentially be read on a certain matrix of distance called the *Euclidean distance matrix* [56].

4.2.2. Gear fault diagnosis

We have a collaboration with Barau (*Safran Tech*) and Hubert (*Safran Tech*), and Dagher (Research Engineer, Inria Chile) on the symbolic-numeric study of the new multi-carrier demodulation method developed in [71]. *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, [71] has proposed to recover each function from the global signal by means of an optimal reconstruction problem which, by means of Fourier analysis, can be rewritten as

$$\text{argmin}_{u\in\mathbb{C}^n, v_1, v_2\in\mathbb{C}^m} \|M - u\, v_1^{\bigstar} - D\, u\, v_2^{\bigstar}\|_F,$$

where $M \in \mathbb{C}^{n\times m}$ (resp. $D \in \mathbb{C}^{n\times n}$) is a given (resp. diagonal) matrix with a special shape, $\|\cdot\|_F$ denotes the Frobenius norm, and $v^{\bigstar}$ the Hermitian transpose of v. Based on closed-form solutions of the exact problem $-$ which are defined by a system of polynomial equations in the unknowns $-$ we have recently proposed efficient numerical algorithms to numerically solve the problem. The first results are interesting and they will be further developed and tested on different data sets. Finally, we shall continue to study the extremal solutions of the corresponding polynomial problem by means of symbolic and numeric methods, etc.

4.3. Applications of the parameter estimation problem to multidisciplinary domains $-$ collaboration with an INSERM team (Rouen University)

For linear systems, the closed-form expressions of the parameters obtained by means of the algebraic parameter estimation problem will continue to provide robust estimates in our multidisciplinary collaborations, in marine biology and human-machine interactions, as it is already the case of existing NON-A results (in collaboration with LOKI, Inria Lille - Nord Europe).

For nonlinear systems, a collaboration with biologists and modelers has been developed for a few years already [29]. Our partners are a team from the Applied Mathematical Department of Le Havre University (modelers) and an INSERM team at Rouen University (neurobiologists). The targeted biological problem is the cortical spreading depression, a brain disease likely to occur after cerebrovascular accidents [60]. We seek $-$ ultimately $-$ a mathematical integro-differential model permitting to predict the triggering of this disease for patients arising in the emergency services of hospitals. The key phenomenon to reproduce is a slow depolarization wave of neurons. Our approach is original because it focuses on the role calcium fluxes in neurons and astrocytes.

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 have already been requested by users long before they become public is a good sign for their wide future impact once in CGAL. 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. We have already referred above to our collaboration with Rien van de Weijgaert [47], [63] (astrophysicist, Groningen, NL). Michael Schindler [59] (theoretical physicist, ENSPCI, CNRS, France) is using our prototype software for 3D periodic weighted triangulations. Stephen Hyde and Vanessa Robins (applied mathematics and physics at Australian National University) have recently signed a software license agreement with INRIA that allows their group to use our prototype 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 [37] 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: http://vegas.loria.fr/qi/.

GRACE Project-Team

4. Application Domains

4.1. Internet of Things

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

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

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

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

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

4.2. Cloud storage

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

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

4.3. Blockchains

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

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

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

LFANT Project-Team (section vide)

OURAGAN 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 progresses 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. In particular, 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, also published in [13]. 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. For example, one can cite different proofs for the 40 possible solutions to the direct kinematics problem for steward platforms and companion experiments based on Gröbner basis computations. 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, for some quite large classes, global algebraic studies allow to propose exhaustive classifications that cannot be reached by other methods.

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: cusps in the workspace).

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 [41] [40] ([58], [60], [59]) 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 *Algorithmic Geometry*.

4.3. Control theory

Many problems in control theory have been studied using general exact polynomial solvers in the past. One can cite the famous Routh-Hurwitz criterion (late 19th century) for the stability of a linear time invariant (LTI) control system and its relation with Sturm sequences and Cauchy index. However most of the strategies used were involving mostly tools for univariate polynomials and then tried to tackle multivariate problems recursively with respect to the variables. More recent work are using a mix of symbolic/numeric strategies, using semi-definite programming for classes of optimization problems or homotopy methods for some algebraic problems, but still very few practical experiments are currently involving certified algebraic using general solvers for polynomial equations.

Our work in control theory is a recent activity and it is done in collaboration with a group of specialists, the GAIA team, Inria Lille-Nord Europe. We started with a well-known problem, the study of the stability of differential delay systems and multidimensional 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, now allow some previously impossible computations and lead to a better understanding of the problems to be solved [37], [36]. The field is porous to computer algebra since one finds for a long time algebraic criteria of all kinds but the technology seems blocked on a recursive use of one-variable methods, whereas our approach involves the direct processing of problems into a larger number of variables or variants.

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 example, we have shown 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 the resolution of bivariate systems.

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.

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

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

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

CAMUS Team

4. Application Domains

4.1. Application Domains

Performance being our main objective, our developments' 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 main 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 Team

4. Application Domains

4.1. Compute-intensive Loop Kernels

The CASH team targets HPC programs, at different levels. Small computation kernels (tens of lines of code) that can be analyzed and optimized aggressively, medium-size kernels (hundreds of lines of code) that require modular analysis, and assembly of compute kernels (either as classical imperative programs or written directly in a dataflow language).

The work on various application domains and categories of programs is driven by the same idea: exploring various topics is a way to converge on unifying representations and algorithms even for specific applications. All these applications share the same research challenge: find a way to integrate computations, data, mapping, and scheduling in a common analysis and compilation framework.

Typical HPC kernels include linear solvers, stencils, matrix factorizations, BLAS kernels, etc. Many kernels can be found in the Polybench/C benchmark suite [46]. The irregular versions can be found in [47]. Numerical kernels used in quantitative finance [57] are also good candidates, e.g., finite difference and Monte-Carlo simulation.

4.2. Medium-size HPC applications

The medium-size applications we target are streaming algorithms [19], scientific workflows [52], and also the now very rich domain of deep learning applications [40]. We explore the possibilities of writing (see Section 3.1) and compiling (see Section 3.3) applications using a dataflow language. As a first step, we will target dataflow programs written in SigmaC [22] for which the fine grain parallelism is not taken into account. In parallel, we will also study the problem of deriving relevant (with respect to safety or optimization) properties on dataflow programs with array iterators.

Obviously, large applications are not limited to assembly of compute kernels. Our languages and formalism definitions (3.1) and analyses (3.2) must also be able to deal with general programs. Our targets also include generalist programs with complex behaviors such as recursive programs operating on arrays, lists and trees; worklist algorithms (lists are not handled within the polyhedral domain). Analysis on these programs should be able to detect non licit memory accesses, memory consumption, hotspots, ..., and to prove functional properties.

The simulation activities (3.5) are both applied internally in CASH, to simulate intermediate representations, and for embedded systems. We are interested in Transaction-Level Models (TLM) of Systems-on-a-Chip (SoCs) including processors and hardware accelerators. TLM provides an abstract but executable model of the chip, with enough details to run the embedded software. We are particularly interested in models written in a loosely timed coding style. We plan to extend these to heterogeneous simulations including a SystemC/TLM part to model the numerical part of the chip, and other simulators to model physical parts of the system.

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. 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. Any computer usage

The PACAP team is working on the fundamental technologies for computer science: processor architecture, performance-oriented compilation and guaranteed response time for real-time. The research results may have 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.

AOSTE2 Team

4. Application Domains

4.1. Avionics

Participants: Liliana Cucu, Keryan Didier, Adriana Gogonel, Cristian Maxim, Dumitru Potop Butucaru, Yves Sorel.

A large number of our activities, in analysis, modelling, design and implementation of real-time embedded systems addresses specific applications mainly in the avionics field (with partners such as Airbus, Thales, Safran, etc.) (in the ASSUME project 9.2.1.1).

4.2. Many-Core Embedded Architectures

Participants: Liliana Cucu, Keryan Didier, Dumitru Potop Butucaru, Yves Sorel.

The AAA approach (fitting embedded applications onto embedded architectures) requires a sufficiently precise description of (a model of) the architecture (description platform). Such platforms become increasingly heterogeneous, and we had to consider a number of emerging ones with that goal in mind, such as Kalray MPPA (in the ASSUME project 9.2.1.1).

4.3. Railways

Participants: Liliana Cucu, Adriana Gogonel, Walid Talaboulma.

The statistical estimation of bounds on the execution time of a program on a processor is applied in the context of railroad crossing in the context of the collaborative project DEPARTS 9.1.2.2 .

HYCOMES Project-Team (section vide)

KAIROS 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

Our local collaborations on handheld, smartphone-like appliances have come to a close with the disappearance of most industrial partners at Sophia Antipolis (Texas Instruments mostly) and the end of the CIM PACA Design platform association. We are renewing collaborations with other local partners, with a focus on Smart Contract applied to connected objects in a IoT environment, and special concern for cloud/fog/edge allocation of computations, expressed with logical time modeling constraints. A speculative european consortium is put up under coordination by Easy Global Market (Sophia-based), and other initiatives with companies such as Symag, Accenture Labs Sophia, and Renault are also being developed.

PARKAS Project-Team (section vide)

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. Industrial Design Tools

The commercially available design tools (such as UML with real-time extensions, MATLAB/ SIMULINK/ dSPACE[0]) and execution platforms (OS such as VXWORKS, QNX, real-time versions of LINUX ...) start now to provide, besides their core functionalities, design or verification methods. Some of them, founded on models of reactive systems, come close to tools with a formal basis, such as for example STATEMATE by iLOGIX.

Regarding the synchronous approach, commercial tools are available: SCADE[0] (based on LUSTRE), CONTROLBUILD and RT-BUILDER (based on SIGNAL) from GEENSYS[0] (part of DASSAULTSYSTEMES), specialized environments like CELLCONTROL for industrial automatism (by the Inria spin-off ATHYS– now part of DASSAULTSYSTEMES). One can observe that behind the variety of actors, there is a real consistency of the synchronous technology, which makes sure that the results of our work related to the synchronous approach are not restricted to some language due to compatibility issues.

4.3. Current Industrial Cooperations

Regarding applications and case studies with industrial end-users of our techniques, we cooperate with Thales on schedulability analysis for evolving or underspecified real-time embedded systems, with Orange Labs on software architecture for cloud services and with Daimler on reduction of nondeterminism and analysis of deadline miss models for the design of automotive systems.

[0] http://www.dspaceinc.com
[0] http://www.esterel-technologies.com
[0] http://www.geensoft.com

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 Signal like Polychrony and developed by TNI, was industrially proven and deployed for that purpose at Peugeot. It served to develop the virtual platform simulating all on-board electronics of PSA cars. This ‘hardware in the loop” simulator was used to test equipments supplied by other manufacturers with respect to virtual cars. In the advent of the related automotive standard, RT-Builder then became AUTOSAR-Builder.

4.2. Factory Automation

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

Model-based analysis in factory automation emerges from different scientific domains and focus on different CPS abstractions that interact in subtle ways: logic of PLC programs, real-time electromechanical processing, physical and chemical environments. This yields domain communication problems that render individual domain analysis useless. For instance, if one domain analysis (e.g. software) modifies a system model in a way that violates assumptions made by another domain (e.g. chemistry) then the detection of its violation may well be impossible to explain to either 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 [30] 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[16]. In a recent article[17] 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 [25].

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 § 6.5) 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, 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)

GALLIUM Project-Team

4. Application Domains

4.1. High-assurance software

A large part of our work on programming languages and tools focuses on improving the reliability of software. Functional programming, program proof, and static type-checking contribute significantly to this goal.

Because of its proximity with mathematical specifications, pure functional programming is well suited to program proof. Moreover, functional programming languages such as OCaml are eminently suitable to develop the code generators and verification tools that participate in the construction and qualification of high-assurance software. Examples include Esterel Technologies's KCG 6 code generator, the Astrée static analyzer, the Caduceus/Jessie program prover, and the Frama-C platform. Our own work on compiler verification combines these two aspects of functional programming: writing a compiler in a pure functional language and mechanically proving its correctness.

Static typing detects programming errors early, prevents a number of common sources of program crashes (null dereferences, out-of bound array accesses, etc), and helps tremendously to enforce the integrity of data structures. Judicious uses of generalized abstract data types (GADTs), phantom types, type abstraction and other encapsulation mechanisms also allow static type checking to enforce program invariants.

4.2. Software security

Static typing is also highly effective at preventing a number of common security attacks, such as buffer overflows, stack smashing, and executing network data as if it were code. Applications developed in a language such as OCaml are therefore inherently more secure than those developed in unsafe languages such as C.

The methods used in designing type systems and establishing their soundness can also deliver static analyses that automatically verify some security policies. Two examples from our past work include Java bytecode verification [38] and enforcement of data confidentiality through type-based inference of information flow and noninterference properties [41].

4.3. Processing of complex structured data

Like most functional languages, OCaml is very well suited to expressing processing and transformations of complex, structured data. It provides concise, high-level declarations for data structures; a very expressive pattern-matching mechanism to destructure data; and compile-time exhaustiveness tests. Therefore, OCaml is an excellent match for applications involving significant amounts of symbolic processing: compilers, program analyzers and theorem provers, but also (and less obviously) distributed collaborative applications, advanced Web applications, financial modeling tools, etc.

4.4. Rapid development

Static typing is often criticized as being verbose (due to the additional type declarations required) and inflexible (due to, for instance, class hierarchies that must be fixed in advance). Its combination with type inference, as in the OCaml language, substantially diminishes the importance of these problems: type inference allows programs to be initially written with few or no type declarations; moreover, the OCaml approach to object-oriented programming completely separates the class inheritance hierarchy from the type compatibility relation. Therefore, the OCaml language is highly suitable for fast prototyping and the gradual evolution of software prototypes into final applications, as advocated by the popular "extreme programming" methodology.

4.5. 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.

MARELLE Project-Team (section vide)

MEXICO Project-Team

4. Application Domains

4.1. Telecommunications

Participants: Stefan Haar, Serge Haddad.

Stefan Haar, Serge Haddad.

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

- In the domain of service oriented computing, it is often necessary to insert some Web service into an existing orchestrated business process, e.g. to replace another component after failures. This requires to ensure, often actively, conformance to the interaction protocol. One therefore needs to synthesize adaptators for every component in order to steer its interaction with the surrounding processes.
- Still in the domain of telecommunications, the supervision of a network tends to move from out-of-band technology, with a fixed dedicated supervision infrastructure, to in-band supervision where the supervision process uses the supervised network itself. This new setting requires to revisit the existing supervision techniques using control and diagnosis tools.

Currently, we have no active cooperation on these subjects.

4.2. Biological Systems

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

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

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

4.3. Autonomous Vehicles

Participant: Serge Haddad.

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

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, even if their capabilities will not allow to reach the so-called quantum supremacy at first.

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 [39]). 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 [27]. 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.) [19]. 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 of application 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. Trustworthy implementations of theorem proving techniques

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

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

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

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

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

- The first one is about formulating automated reasoning techniques in terms of inference systems, so that they fit the approach described above. While this is rather standard for technique used in first-order Automated Theorem Provers (ATP), such as resolution, superposition, etc, this is much less standard in SMT-solving, the branch of automated reasoning that can natively handle reasoning in a combination of mathematical theories: the traditional techniques developed there usually organise the collaborations between different reasoning black boxes, whose opaque mechanisms less clearly connect to proof-theoretical inference systems. We are therefore investigating new foundations for reasoning in combinations of theories, expressed as fine-grained inference systems, and developed the *Conflict-Driven Satisfiability framework* for these foundations [13].
- The second one is about understanding how to deal with quantifiers in presence of one or more theories: On the one hand, traditional techniques for quantified problems, such as *unification* [29] or *quantifier elimination* are usually designed for either the empty theory or very specific theories. On the other hand, the industrial techniques for combining theories (Nelson-Oppen, Shostak, MCSAT [64], [69], [73], [53]) are designed for quantifier-free problems, and quantifiers there are dealt with incomplete *clause instantiation* methods or *trigger*-based techniques [41]. We are working on making the two approaches compatible.
- The above architecture's modular approach raises the question of how its different modules can safely cooperate (in terms of guaranteed correctness), while some of them are trusted and others are not. The issue is particularly acute if some of the techniques are run concurrently and exchange data at unpredictable times. For this we explore new solutions based on Milner's *LCF* [63]. In [47], we argued that our solutions in particular provide a way to fulfil the "Strategy Challenge for SMT-solving" set by De Moura and Passmore [74].

4.2. Principled computation for strong lambda-calculi

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

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

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

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

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

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

PI.R2 Project-Team (section vide)

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.

4.4. Systems Biology

Systems Biology is a recent topic in SUMO. In systems biology, many continuous variables interact together. Biological systems are thus good representatives for large complex quantitative systems, for which we are developing analysis and management methods. For instance, the biological pathway of apoptosis explains how numerous molecules interact inside a cell, triggered by some outside signal (drug, etc.), eventually leading to the death of the cell by apoptosis. While intrinsically quantitative in nature and in problems, data are usually noisy and problems need not be answered with ultimate precision. It thus seems reasonable to resort to approximations in order to handle the state-space explosion resulting from the high dimensionality of biological systems.

We are developing models and abstraction tools for systems biology. Studying these models suggests new reduction methods, such as considering populations instead of explicitly representing every single element into play (be it cells, molecules, etc): we thus develop algorithms handling a population symbolically, either in a continuous (probability distribution) or a discrete (parametric) way. An intermediate goal is to speed-up the analysis of such systems using abstractions, and a long term goal is to develop top-down model-checking methods that can be run on these abstractions.

4.5. Formal Verification of Smart Flexible Manufacturing Systems

Modern production/assembly lines are based on generic multipurpose programmable tools that are quickly reassembled and reprogrammed to accomodate new production processes. In a similar manner, complex products are also reengineered by assembling existing elementary functions, together with their correponding software. This modular construction principle enables a fast redesign of products or assembly chains, at the expense of possibly introducing bugs or malfunctions. Verification is thus a crucial step to guarantee the correctness of these systems. In particular, timing aspects are essential in order to both check correctness of an assembling with respect to some specification, but also in order to design software sensors that help the online monitoring of a system. The main challenges here essentially lie in the selection of appropriate verification formalisms, in the derivation of models for the systems under study, and in the size of the systems to handle.

TOCCATA Project-Team

4. Application Domains

4.1. Domain 1

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

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

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

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

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

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

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

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.

Formal verification techniques can contribute to certifying the correctness of systems. In particular, they help assert under which assumptions an algorithm or system functions as required. For example, the highest levels of the Common Criteria for Information Technology Security Evaluation encourage the use of formal methods. While initially the requirements of certified development were mostly restricted to safety-critical systems, the cost of unavailable services due to malfunctioning system components and software provides wider incentives for verification. For example, we have been working on modeling and verifying medical devices that require closed-loop models of both the system and its environment.

CIDRE Project-Team

4. Application Domains

4.1. Security is required everywhere

With the infiltration of computers and software in almost all aspects of our modern life, security can nowadays be seen as an absolutely general concern. As such, the results of the research targeted by CIDRE apply to a wide range of domains. It is clear that critical systems, in which security (and safety) is a major concern, can benefit from ideas such as dynamic security policy monitoring. On the other hand, systems used by the general public (basically, the internet and services such as web or cloud services, social networks, location-based services, etc.) can also benefit from the results obtained by CIDRE, in particular to solve some of the privacy issues raised by these systems that manipulate huge amounts of personal data. In addition, systems are getting more and more complex, decentralized, distributed, or spontaneous. Cloud computing, in particular, brings many challenges that could benefit from ideas, approaches and solutions studied by CIDRE in the context of distributed systems.

Industrial Control Systems (ICS) and in particular Supervisory Control and Data Acquisition are also new application domains for intrusion detection. The Stuxnet attack has emphasized the vulnerability of such critical systems which are not totally isolated anymore. Securing ICS is challenging since modifications of the systems, for example to patch them, are often not possible. High availability requirements also often conflict with preventive approaches. In this case, security monitoring is appealing to protect such systems against malicious activities. Intrusion detection in ICS is not fundamentally different from traditional approaches. However, new hypotheses and constraints need to be taken into account, which also bring interesting new research challenges.

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. CyberStrategy/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 or ssh, 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

4. Application Domains

4.1. System analysis

The work performed in Axes 1 and 2 and the methods developed there are applicable to the domain of system analysis, both wrt. program analysis and hardware analysis.

4.2. Cybersecurity

The work done in the axes above aims at improving cybersecurity, be it via vulnerability analyses, malware analyses and the development of safer networking mechanisms.