RESEARCH CENTER

FIELD
Algorithmics, Programming, Software and Architecture

Activity Report 2012

Section Application Domains

Edition: 2013-04-24

ARIC Team

4. Application Domains

4.1. Hardware Arithmetic

The application domains of hardware arithmetic operators are digital signal processing, image processing, embedded applications, reconfigurable computing, and cryptography.

4.2. Floating-point and Validated Numerics

Our expertise on validated numerics is useful to analyse and improve, and guarantee the quality of numerical results in a wide range of applications, from scientific simulation to global optimization or control theory. Much of our work, in particular the development of correctly rounded elementary functions, is critical to the reproducibility of floating-point computations.

4.3. Cryptography, Cryptology, Communication Theory

Lattice reduction algorithms have direct applications in public-key cryptography. A new and promising field of applications is communications theory.

CARAMEL Project-Team

4. Application Domains

4.1. Cryptology

The first application domain for our research is cryptology. This includes cryptography (constructive side) and cryptanalysis (breaking systems). For the cryptanalysis part, although it has practical implications, we do not expect any transfer in the classical sense of the term: it is more directed to governmental agencies and the end-users who build their trust, based on the cryptanalysis effort.

4.1.1. Cryptography

Our cryptographic contributions are related to multiple facets of the large realm of curve-based cryptology. While it is quite clear that enough algorithms exist in order to provide cryptographers with elliptic curves having a suitably hard discrete logarithm (as found in cryptographic standards for instance), one must bear in mind that refinements of the requirements and extensions to curves of higher genus raise several interesting problems. Our work contributes to expanding the cryptographer's capabilities in these areas.

In the context of genus-2 curves, our work aims at two goals. First, improvements on the group law on selected curves yield better speed for the associated cryptosystems. The cryptographic primitives, and then the whole suite of cryptographic protocols built upon such curves would be accelerated. The second goal is the expansion of the set of curves that can be built given a set of desired properties. Using point counting algorithms for arbitrary curves, a curve offering a 128-bit security level, together with nice properties for fast arithmetic, has been computed by CARAMEL [3]. Another natural target for construction of curves for cryptography is also the suitability of curves for pairings. We expect to be able to compute such curves.

Implementations of curve-based cryptography, both in hardware and software, are a necessary step on the way to assessing cryptographic speed. We plan to provide such implementations. In particular, on the hardware side, one of our goals is the design of a complete cryptographic coprocessor, including all the primitives for curve-based and pairing-based cryptography, providing optimized and configurable efficiency vs area trade-off.

4.1.2. Cryptanalysis

Our research on cryptanalysis is important for the cryptographic industry: by detecting weak instances, and setting new records we contribute to the definition of recommended families of systems together with their key sizes. The user's confidence in a cryptographic primitive is also related to how well the underlying problem is studied by researchers.

In particular, our involvement in computations with "NFS-like" algorithms encompasses of course the task of assessing the computational limits for integer factorization and discrete-logarithm computations. The impact of the former is quite clear as it concerns the RSA algorithm; record-sized computations attract broad interest and determine updates on key-length recommendations. The latter are particularly important for pairing-based cryptography, since, in this context, one naturally encounters discrete-logarithm problems in extension fields of large degree.

4.2. Standardization

4.2.1. Floating-point arithmetic

The IEEE 754 standard for floating-point arithmetic was revised in 2008. The main new features are some new formats for decimal computations, and the recommendation of correctly rounded transcendental functions. The new decimal formats should not have an impact on our work, since we either use integer-only arithmetic, or arbitrary-precision binary floating-point arithmetic through the GNU MPFR library.

A new standard (P1788) is currently under construction for interval arithmetic. We are not officially involved in this standard, but we follow the discussions, to check in particular that the proposed standard will also cover arbitrary precision (interval) arithmetic.

4.3. Computer algebra systems

Some of our software libraries are being used by computer algebra systems. Most of those libraries are free software, with a license that allows proprietary systems to link them. This gives us a maximal visibility, with a large number of users.

4.3.1. Magma

Magma is a very large computational algebra package. It provides a mathematically rigorous environment for computing with algebraic, number-theoretic, combinatoric, and geometric objects. It is developed in Sydney, by the team around John Cannon. It is non-commercial (in the sense that its goal is not to make profit), but is not freely distributed and is not open-source.

Several members of the team have visited Sydney — a few years ago — to contribute to the development of Magma, by implementing their algorithms or helping in integrating their software. Our link to Magma exists also via the libraries it uses: it currently links GNU MPFR and GNU MPC for its floating-point calculations, and links GMP-ECM as part of its factorization suite.

4.3.2. Pari-GP

Pari/GP is a computational number theory system that is composed of a C library and an interpreter on top of it. It is developed in Bordeaux, where Karim Belabas from the LFANT project-team is the main maintainer. Its license is GPL. Although we do not directly contribute to this package, we have good contact with the developers.

4.3.3. Sage

Sage is a fairly large scale and open-source computer algebra system written in Python. Sage aggregates a large amount of existing free software, aiming at the goal of selecting the fastest free software package for each given task. The motto of Sage is that instead of “reinventing the wheel” all the time, Sage is “building the car”. To date, Sage links GNU MPFR, GMP-ECM, and GNU MPC as standard packages.

CASCADE Project-Team

4. Application Domains

4.1. Hash Functions

Since the previous section just ended on this topic, we start with it for the major problems to address within the next years. A NIST competition on hash functions has been launched late 2007 and finished a few months ago. In the first step, cryptographers had to build and analyze their own candidate; in a second step, cryptanalysts were solicited, in order to analyze and break all the proposals. The conclusion has been announced with the winner Keccak, on October 2nd, 2012.

The symmetric people of the Cascade team have worked these years on the development of a new hash function called SIMD that has been selected for the second round of the NIST SHA-3 competition. SIMD hash function is quite similar to members of the MD/SHA family. It is based on a familiar Merkle-Damgard design, where the compression function is built from a Feistel-like cipher in Davies-Meyer mode. However there are some innovations in this design: the internal state is twice as big as the output size, we use a strong message expansion, and we use a modified feed-forward in the compression function. The main design criteria was to follow the MD/SHA designs principle which are quite well understood, and to add some elements to avoid all known attacks. SIMD is particularly efficient on platforms with vector instructions (SIMD) which are available on many processors. Such instructions have been proposed since 1997 and are now widely deployed. Moreover, it is also possible to use two cores on multicore processors to boost the performances with a factor 1.8 by splitting the message expansion function and the hashing process.

More recently, we essentially worked on the other candidates, with some analyses and attacks. Even if the winner has been selected, there is still a lot of work to do on hash functions, as there is on block-ciphers, even if AES was selected a long time ago.

4.2. Anonymity and Privacy

A relatively new goal of growing importance of cryptography is *privacy*. In a digital world where data is ubiquitous, users are more and more concerned about confidentiality of their personal data. Cryptography makes it possible to benefit from the advantages of digital technology while at the same time providing means for privacy protection. An example is anonymous authentication: A user can convincingly prove that she has certain rights without however revealing her identity. Privacy and anonymity remains thus one of the main challenges for the next years.

4.3. Copyright Protection

Similarly to the privacy concern, the digital world makes easy the large-scale diffusion of information. But in some cases, this can be used in violation of some copyrights. Cryptography should help at solving this problem, which is actually two-fold: one can either mark the original document in order to be able to follow the distribution (and possibly trace the traitor who illegally made it public) or one can publish information in an encrypted way, so that authorized people only can access it.

4.4. Lattice-Based Cryptography

In 1996, Ajtai [66] showed that lattices, which up to that point had only been used as tools in cryptanalysis, can actually be used to *construct* cryptographic primitives. He proposed a cryptographic primitive whose security is based on the worst-case hardness of lattice problems: if one succeeds in breaking the primitive, even with some small probability, then one can also solve any instance of a certain lattice problem. This powerful property makes lattice-based cryptographic constructions very attractive. In contrast, virtually all other cryptographic constructions are based on some average-case assumption. Furthermore, there are currently

very few alternatives to traditional number-theoretic based cryptography such as RSA. Such alternatives will be needed in case an efficient algorithm for factoring integers is ever found, a possibility some leading number theorists consider as quite likely. In fact, efficient quantum algorithms for factoring integers and computing discrete logarithms already exist [87]. Although large-scale quantum computers are not expected to exist for at least a decade, this fact should already be regarded as a warning. In contrast, there are currently no known quantum algorithms for lattice problems. Finally, the computations involved in lattice-based cryptography are typically very fast and often require only modular additions, making them attractive for many applications.

For all these reasons, lattice-based cryptography has become a hot topic, especially in the last few years, and our group is playing an important part in this effort.

4.5. Cryptanalysis

As already explained, even with the *provable security* concept, cryptanalysis is still an important area, and attacks can be done at several levels. Algebraic tools (against integer factoring, discrete logarithm, polynomial multivariate systems, lattice reduction, etc) have thus to be studied and improved in order to further evaluation of the actual security level of cryptographic schemes.

At the hardware level, side-channel information has to be identified (time, power, radiation, noise, heat, etc) in order to securely protect embedded systems. But such information may also be used in a positive way....

4.6. Access Control

Access control policies describe the rights that different users have on the resources available and are used to protect systems against attacks by unauthorised users. Many authorisation and access control models have been proposed so far; we focus mainly on category-based access control. Amongst the most important issues that have to be addressed in this area are the development of techniques and tools for the analysis and verification of access control policies, and the design of transparent and efficient mechanisms to enforce access control policies.

GALAAD Project-Team

4. Application Domains

4.1. Shape modeling

Geometric modeling is increasingly familiar for us (synthesized images, structures, vision by computer, Internet, ...). Nowadays, many manufactured objects are entirely designed and built by means of geometric software which describe with accuracy the shape of these objects. The involved mathematical models used to represent these shapes have often an algebraic nature. Their treatment can be very complicated, for example requiring the computations of intersections or isosurfaces (CSG, digital simulations, ...), the detection of singularities, the analysis of the topology, etc. Optimising these shapes with respect to some physical constraints is another example where the choice of the models and the design process are important to lead to interesting problems in algebraic geometric modeling and computing. We propose the developments of methods for shape modeling that take into account the algebraic specificities of these problems. We tackle questions whose answer strongly depends on the context of the application being considered, in direct relationship with the industrial contacts that we are developing in Computer Aided Geometric Design.

4.2. Shape processing

Many problems encountered in the application of computer sciences start from measurement data, from which one wants to recover a curve, a surface, or more generally a shape. This is typically the case in image processing, computer vision or signal processing. This also appears in computer biology where the geometry of distances plays a significant role, for example, in the reconstruction from NMR (Nuclear Magnetic Resonance) experiments, or the analysis of realizable or accessible configurations. In another domain, scanners which tend to be more and more easily used yield large set of data points from which one has to recover compact geometric model. We are working in collaboration with groups in agronomy on the problems of reconstruction of branching models (which represent trees or plants). We are investigating the application of algebraic techniques to these reconstruction problems. Geometry is also highly involved in the numerical simulation of physical problems such as heat conduction, ship hull design, blades and turbines analysis, mechanical stress analysis. We apply our algebraic-geometric techniques in the isogeometric approach which uses the same (bspline) formalism to represent both the geometry and the solutions of partial differential equations on this geometry.

GEOMETRICA Project-Team

4. Application Domains

4.1. Geometric Modeling and Shape Reconstruction

Modeling 3D shapes is required for all visualization applications where interactivity is a key feature since the observer can change the viewpoint and get an immediate feedback. This interactivity enhances the descriptive power of the medium significantly. For example, visualization of complex molecules helps drug designers to understand their structure. Multimedia applications also involve interactive visualization and include e-commerce (companies can present their products realistically), 3D games, animation and special effects in motion pictures. The uses of geometric modeling also cover the spectrum of engineering, computer-aided design and manufacture applications (CAD/CAM). More and more stages of the industrial development and production pipeline are now performed by simulation, due to the increased performance of numerical simulation packages. Geometric modeling therefore plays an increasingly important role in this area. Another emerging application of geometric modeling with high impact is medical visualization and simulation.

In a broad sense, shape reconstruction consists of creating digital models of real objects from points. Example application areas where such a process is involved are Computer Aided Geometric Design (making a car model from a clay mockup), medical imaging (reconstructing an organ from medical data), geology (modeling underground strata from seismic data), or cultural heritage projects (making models of ancient and or fragile models or places). The availability of accurate and fast scanning devices has also made the reproduction of real objects more effective such that additional fields of applications are coming into reach. The members of GEOMETRICA have a long experience in shape reconstruction and contributed several original methods based upon the Delaunay and Voronoi diagrams.

4.2. Scientific Computing

Meshes are the basic tools for scientific computing using finite element methods. Unstructured meshes are used to discretize domains bounded by complex shapes while allowing local refinements. GEOMETRICA contributes to mesh generation of 2D and 3D possibly curved domains. Most of our methods are based upon Delaunay triangulations, Voronoi diagrams and their variants. Anisotropic meshes are also investigated. We investigate in parallel both greedy and variational mesh generation techniques. The greedy algorithms consist of inserting vertices in an initial coarse mesh using the Delaunay refinement paradigm, while the variational algorithms consists of minimizing an energy related to the shape and size of the elements. Our goal is to show the complementarity of these two paradigms. Quadrangle surface meshes are also of interest for reverse engineering and geometry processing applications. Our goal is to control the final edge alignment, the mesh sizing and the regularity of the quadrangle tiling.

GRACE Team

3. Application Domains

3.1. Cryptology

We want to establish the security of practical proposals relying on computational problems, be they standardized (like RSA or Elliptic Curve Cryptography), or more exotic (like Hyperelliptic Curve Cryptography). We do not work with abstract cryptographic primitives. On the design side, building efficient near-optimal codes impacts directly on the security of basic operations in symmetric primitives. We also investigate other applications, such as secret sharing schemes, universal hash functions, and message authentication, revisiting them in the context of Algebraic Geometry codes.

3.2. Codes in Computer Science

We do not want to do basic forward error correction, dealing with bit error rates and signal-to-noise ratios. Rather, we aim to deal with higher models of communication and computation, including peer-to-peer systems and distributed storage. We also consider adversarial noise, or distributed computations with byzantine faults. List decoding deals precisely with these kinds of “difficult”, non-random errors. In a related spirit, one can deal with “computationally bounded channels”, where the errors are generated by an adversarial machine or algorithm that is computationally bounded.

LFANT Project-Team

4. Application Domains

4.1. Number theory

Being able to compute quickly and reliably algebraic invariants is an invaluable aid to mathematicians: It fosters new conjectures, and often shoots down the too optimistic ones. Moreover, a large body of theoretical results in algebraic number theory has an asymptotic nature and only applies for large enough inputs; mechanised computations (preferably producing independently verifiable certificates) are often necessary to finish proofs.

For instance, many Diophantine problems reduce to a set of Thue equations of the form $P(x, y) = a$ for an irreducible, homogeneous $P \in \mathbb{Z}[x, y]$, $a \in \mathbb{Z}$, in unknown integers x, y. In principle, there is an algorithm to solve the latter, provided the class group and units of a rupture field of P are known. Since there is no other way to prove that the full set of solutions is obtained, these algebraic invariants must be computed and certified, preferably without using the GRH.

Deeper invariants such as the Euclidean spectrum are related to more theoretical concerns, e.g., determining new examples of principal, but not norm-Euclidean number fields, but could also yield practical new algorithms: Even if a number field has class number larger than 1 (in particular, it is not norm-Euclidean), knowing the upper part of the spectrum should give a *partial* gcd algorithm, succeeding for almost all pairs of elements of $\mathcal{O}_K$. As a matter of fact, every number field which is not a complex multiplication field and whose unit group has rank strictly greater than 1 is almost norm-Euclidean [37], [38].

Algorithms developed by the team are implemented in the free PARI/GP system for number theory maintained by K. Belabas, which is a reference and the tool of choice for the worldwide number theory community.

4.2. Cryptology

Public key cryptology has become a major application domain for algorithmic number theory. This is already true for the ubiquitous RSA system, but even more so for cryptosystems relying on the discrete logarithm problem in algebraic curves over finite fields [7]. For the same level of security, the latter require smaller key lengths than RSA, which results in a gain of bandwidth and (depending on the precise application) processing time. Especially in environments that are constrained with respect to space and computing power such as smart cards and embedded devices, algebraic curve cryptography has become the technology of choice. Most of the research topics of the LFANT team concern directly problems relevant for curve-based cryptology: The difficulty of the discrete logarithm problem in algebraic curves determines the security of the corresponding cryptosystems. Complex multiplication, point counting and isogenies provide, on one hand, the tools needed to create secure instances of curves. On the other hand, isogenies have been found to have direct cryptographic applications to hash functions [39] and encryption [43]. Pairings in algebraic curves have proved to be a rich source for novel cryptographic primitives. Class groups of number fields also enter the game as candidates for algebraic groups in which cryptosystems can be implemented. However, breaking these systems by computing discrete logarithms has proved to be easier than in algebraic curves; we intend to pursue this cryptanalytic strand of research.

Apart from solving specific problems related to cryptology, number theoretic expertise is vital to provide cryptologic advice to industrial partners in joint projects. It is to be expected that continuing pervasiveness and ubiquity of very low power computing devices will render the need for algebraic curve cryptography more pressing in coming years.

POLSYS Project-Team

4. Application Domains

4.1. Cryptology

We propose to develop a systematic use of *structured systems* in Algebraic Cryptanalysis. We want to improve the efficiency and to predict the theoretical complexity of such attacks. We plan to demonstrate the power of algebraic techniques in new areas of cryptography such as Algebraic Number Theory (typically, in curve based cryptography).

4.2. Engineering sciences

Solving polynomial systems over the reals arise as a critical issue in wide range of problems coming from engineering sciences (biology, physics, control theory, etc.). We will focus on developing general enough software that may impact on these domains with a particular focus on control theory.

SECRET Project-Team

4. Application Domains

4.1. Application domains

Our main application domains are:

- cryptology,
- error-correcting codes, especially codes for quantum communications and fault-tolerant quantum computing,
- reverse-engineering of communication systems.

We also investigate some cross-disciplinary domains, which require a scientific competence coming from other areas, mainly social aspects of cryptology, cryptology for large databases.

VEGAS Project-Team

3. Application Domains

3.1. Computer graphics

We are interested in the application of our work to virtual prototyping, which refers to the many steps required for the creation of a realistic virtual representation from a CAD/CAM model.

When designing an automobile, detailed physical mockups of the interior are built to study the design and evaluate human factors and ergonomic issues. These hand-made prototypes are costly, time consuming, and difficult to modify. To shorten the design cycle and improve interactivity and reliability, realistic rendering and immersive virtual reality provide an effective alternative. A virtual prototype can replace a physical mockup for the analysis of such design aspects as visibility of instruments and mirrors, reachability and accessibility, and aesthetics and appeal.

Virtual prototyping encompasses most of our work on effective geometric computing. In particular, our work on 3D visibility should have fruitful applications in this domain. As already explained, meshing objects of the scene along the main discontinuities of the visibility function can have a dramatic impact on the realism of the simulations.

3.2. Solid modeling

Solid modeling, i.e., the computer representation and manipulation of 3D shapes, has historically developed somewhat in parallel to computational geometry. Both communities are concerned with geometric algorithms and deal with many of the same issues. But while the computational geometry community has been mathematically inclined and essentially concerned with linear objects, solid modeling has traditionally had closer ties to industry and has been more concerned with curved surfaces.

Clearly, there is considerable potential for interaction between the two fields. Standing somewhere in the middle, our project has a lot to offer. Among the geometric questions related to solid modeling that are of interest to us, let us mention: the description of geometric shapes, the representation of solids, the conversion between different representations, data structures for graphical rendering of models and robustness of geometric computations.

3.3. Fast prototyping

We work in collaboration with CIRTES on rapid prototyping. CIRTES, a company based in Saint-Dié-des-Vosges, has designed a technique called Stratoconception® where a prototype of a 3D computer model is constructed by first decomposing the model into layers and then manufacturing separately each layer, typically out of wood of standard thickness (e.g. 1 cm), with a three-axis CNC (Computer Numerical Controls) milling machine. The layers are then assembled together to form the object. The Stratoconception® technique is cheap and allows fast prototyping of large models.

When the model is complex, for example an art sculpture, some parts of the models may be inaccessible to the milling machine. These inaccessible regions are sanded out by hand in a post-processing phase. This phase is very consuming in time and resources. We work on minimizing the amount of work to be done in this last phase by improving the algorithmic techniques for decomposing the model into layers, that is, finding a direction of slicing and a position of the first layer.

ALF Project-Team

4. Application Domains

4.1. Application Domains

Performance, processor architecture, compilers,telecommunications, multimedia, biology, health, engineering, environment, transportation

The ALF team is working on the fundamental technologies for computer science: processor architecture and performance-oriented compilation. The research results have impacts on any application domain that requires high performance executions (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.

CAIRN Project-Team

4. Application Domains

4.1. Panorama

keywords: telecommunications, wireless communications, wireless sensor networks, content-based image retrieval, video coding, intelligent transportation systems, automotive, security

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.

The high complexity of the **Next-Generation (4G) Wireless Communication Systems** leads to the design of real-time high-performance specific architectures. The study of these techniques is one of the main field of applications for our research, based on our experience on WCDMA for 3G implementation.

In **Wireless Sensor Networks** (WSN), where each wireless node has to operate without battery replacement for a long time, energy consumption is the most important constraint. In this domain, we mainly study energy-efficient architectures and wireless cooperative techniques for WSN.

Intelligent Transportation Systems (ITS), and especially Automotive Systems, more and more apply technology advances. While wireless transmissions allow a car to communicate with another or even with road infrastructure, **automotive industry** can also propose driver assistance and more secure vehicles thanks to improvements in computation accuracy for embedded systems.

Other important fields will also be considered: hardware cryptographic and security modules, specialized hardware systems for the filtering of the network traffic at high-speed, high-speed true-random number generation for security, content-based image retrieval and video processing.

4.2. 4G Wireless Communication Systems

With the advent of the next generation (4G) broadband wireless communications, the combination of MIMO (Multiple-Input Multiple-Output) wireless technology with Multi-Carrier CDMA (MC-CDMA) has been recognized as one of the most promising techniques to support high data rate and high performance. Moreover, future mobile devices will have to propose interoperability between wireless communication standards (4G, WiMax ...) and then implement MIMO pre-coding, already used by WiMax standard. Finally, in order to maximize mobile devices lifetime and guarantee quality of services to consumers, 4G systems will certainly use cooperative MIMO schemes or MIMO relays. Our research activity focuses on MIMO pre-coding and MIMO cooperative communications with the aim of algorithmic optimization and implementation prototyping.

4.3. Wireless Sensor Networks

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. Cross-layer optimizations lead to energy-efficient architectures and cooperative techniques dedicated to sensor networks applications. In particular, cooperative MIMO techniques are used to decrease the energy consumption of the communications.

4.4. Multimedia processing

In multimedia applications, audio and video processing is the major challenge embedded systems have to face. It is computationally intensive with power requirements to meet. Video or image processing at pixel level, like image filtering, edge detection and pixel correlation or at block-level such as transforms, quantization, entropy coding and motion estimation have to be accelerated. We investigate the potential of reconfigurable architectures for the design of efficient and flexible accelerators in the context of multimedia applications.

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 prior objective, we can expect some positive effects for the following reasons:

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

COMPSYS Project-Team

4. Application Domains

4.1. Compilers for Embedded Computing Systems

The previous sections described our main activities in terms of research directions, but also places Compsys within the embedded computing systems domain, especially in Europe. We will therefore not come back here to the importance, for industry, of compilation and embedded computing systems design.

In terms of application domain, the embedded computing systems we consider are mostly used for multimedia: phones, TV sets, game platforms, etc. But, more than the final applications developed as programs, our main application is the computer itself: how the system is organized (architecture) and designed, how it is programmed (software), how programs are mapped to it (compilation and high-level synthesis).

The industry that can be impacted by our research is thus all the companies that develop embedded systems and processors, and those (the same plus other) than need software tools to map applications to these platforms, i.e., that need to use or even develop programming languages, program optimization techniques, compilers, operating systems. Compsys do not focus on all these critical parts, but our activities are connected to them.

AOSTE Project-Team

4. Application Domains

4.1. Multicore System-on-Chip design

Synchronous formalisms and GALS or multiclock extensions are natural model representations of hardware circuits at various abstraction levels. They may compete with HDLs (Hardware Description Languages) at RTL and even TLM levels. The main originality of languages built upon these models is to be based on formal *synthesis* semantics, rather than mere simulation forms.

The flexibility in formal Models of Computation and Communication allows specification of modular Latency-Insensitive Designs, where the interconnect structure is built up and optimized around existing IP components, respecting some mandatory computation and communication latencies prescribed by the system architect. This allows a real platform view development, with component reuse and timing-closure analysis. The design and optimization of interconnect fabric around IP blocks transform at modeling level an (untimed) asynchronous versions into a (scheduled) multiclock timed one.

Also, Network on Chip design may call for computable switching patterns, just like computable scheduling patterns were used in (predictable) Latency-Insensitive Design. Here again formal models, such as Cyclo-static dataflow graphs and extended Kahn networks with explicit routing schemes, are modeling elements of choice for a real synthesis/optimization approach to the design of systems.

Multicore embedded architecture platform may be represented as Marte UML component diagrams. The semantics of concurrent applications may also be represented as Marte behavior diagrams embodying precise MoCCs. Optimized compilations/syntheses rely on specific algorithms, and are represented as model transformations and allocation (of application onto architecture).

Our current work aims thus primarily at providing Theoretical Computer Science foundations to this domain of multicore embedded SoCs, with possibly efficient application in modeling, analysis and compilation wherever possible due to some natural assumptions. We also deal with a comparative view of Esterel and SystemC TLM for more practical modeling, and the relation between the Spirit IP-Xact interface standard in SoC domain with its Marte counterpart.

4.2. Automotive and avionic embedded systems

Model-Driven Engineering is in general well accepted in the transportation domains, where design of digital software and electronic parts in usually tighly coupled with larger apsects of system design, where models from physics are being used already. The formalisms AADL (for avionics) and AutoSar [55] (for automotive) are providing support for this, unfortunately not always with a clean and formal semantics. Thus there is a strong need here for approaches that bring closer together formal methods and tools on the one hand, engineering best practices on the other hand.

From a structural point of view AUTOSAR succeeded in establishing a framework that provides significant confidence in the proper integration of software components from a variety of distinct suppliers. But beyond those structural (interface) aspects, dynamic and temporal views are becoming more of a concern, so that AUTOSAR has introduced the AUTOSAR Specification of Timing Extension. AUTOSAR (discrete) timing models consist of timing descriptions, expressed by events and event chains, and timing constraints that are imposed on these events and event chains.

An important issue in all such formalisms is to mix in a single design framework heterogeneous time models and tasks: based on different timebases, with different triggering policy (event-triggered and time-triggered), and periodic and/or aperiodic tasks, with distinct periodicity if ever. Adequate modeling is a prerequisite to the process of scheduling and allocating such tasks onto complex embedded architectural platforms (see AAA approach in foundation section 3.3). Only then can one devise powerful synthesis/analysis/verification techniques to guide designers towards optimized solutions.

Traceability is also an important concern, to close the gap between early requirements and constraints modelling on the one hand, verification and correct implementation of these constraints at the different levels of the development on the other hand.

CONVECS 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,
- *Consumer electronics:* home networking, video on-demand,
- *Databases:* transaction protocols, distributed knowledge bases, stock management,
- *Distributed systems:* virtual shared memory, distributed file systems, election algorithms, dynamic reconfiguration algorithms, fault tolerance algorithms, cloud computing,
- *Embedded systems:* smart-card applications, air traffic control, avionic systems,
- *Hardware architectures:* asynchronous circuits, multiprocessor architectures, systems on chip, networks on chip, bus arbitration protocols, 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.

DART Project-Team

4. Application Domains

4.1. Gaspard2 for avionic hybrid test platform design

The emergence and the maturity of FPGA circuits for distributed and reconfigurable architectures offer the opportunity to explore real time problems in the field of avionic systems. FPGA becomes de facto a major processing element as same as general CPUs. As of now, the FPGA is widely used in the field of I/O component in order to connect the real equipment with the CPU host. Among the main features mapped into the FPGA in the original architecture, we quote the fast serial link and RAM IPs (Intellectual property) which are needed to ensure communication between CPU and FPGA. Additionally, the Base Time IP is needed for the global system synchronization. This minimal configuration based on FPGA can be duplicated several times and connected together to build bigger test system or a complete simulator. Eurocopter expectation for the above-described architecture is to prototype some models which can be eligible and relocated in the FPGA. The objective is to increase the performances of these models and to reduce the communication latencies by the means of embedding the different parts in the same chip. To do so, we studied in this first year a real avionic test loop in order to extract the complex models that will be implemented in the FPGA. Different hardware model configurations have been explored to reach an optimal well-balanced global system using the ML403 Virtex-4 Xilinx board. Different tradeoffs in terms of performance and resource occupation in the FPGA are obtained. Later, these results will be used for dynamically adapt the system functioning according to the available resources and performance requirements.

As a second part, we used the MARTE profile to represent an hybrid system (CPU/FPGA).In the MARTE specification, an application is a set of tasks connected through ports. Tasks are considered as mathematical functions reading data from their input ports and writing data on their output ports. This specification has been used to model the avionic test loop. In addition, MARTE allows describing the hardware architecture in a structural way. Typical components such as HwProcessor, HwFPGA and HwRAM can be specified with their non-functional properties. We used this subset of MARTE in order to represent an hybrid multiprocessor architecture. The main component of this architecture is composed of the Xeon-X3370 processor (multicore CPU) and the Virtex-4 Xilinx FPGA. Furthermore, MARTE provides the Allocate concept as well as the concept specially crafted for repetitive structures Distribute. This latter concept gives a way to express regular distribution of tasks onto a set of processors or FPGA resources. The mapping step relies on two types of distribution (timeScheduling and spatialDistibution) depending on the target hardware platform (CPU/FPGA). The different models of our avionic test loop can be mapped onto the host multicore processor, the embedded processor (Microblaze) or the hardware resources in the FPGA.

4.2. Electromagnetic modeling

We collaborate with the L2EP specialized in electromagnetic modeling, on algorithms definition and the parallelization of their computations, especially on GPUs.

For the first point, we have designed a parallel version of the Finite Integration Technique (F.I.T). This is used to simulate electromagnetic phenomena. This technique is efficient if the mesh is generated by a regular hexahedron. Moreover the matrix system, obtained from a regular mesh can be exploited to use the parallel direct solver. In fact, in reordering the unknowns by the nested dissection method, it is possible to construct directly the lower triangular matrix with many processors without assembling the matrix system. During this year, we have used our parallel direct solver as a preconditionner for a sparse linear system coming from a FEM problem with a good efficiency[25].

For the second point, we have include our Gaspard2 generated code in Code_CARMEL, a software for electromagnetic fields simulations. This GPGPU code is now robust enough to run most of the testbenches implemented inside this framework[23].

ESPRESSO Project-Team

4. Application Domains

4.1. Embedded systems

The application domains covered by the Polychrony toolbox are engineering areas where a system design-flow requires high-level model transformations and verifications to be applied during the development-cycle. The project-team has focused on developing such integrated design methods in the context of avionics applications, through the European IST projects Sacres, Syrf, Safeair, Speeds, and through the national ANR projects Topcased, OpenEmbeDD, Spacify. In this context, Polychrony is seen as a platform on which the architecture of an embedded system can be specified from the earliest design stages until the late deployment stages through a number of formally verifiable design refinements.

Along the way, the project adopted the policy proposed with project Topcased and continued with OpenEmbeDD to make its developments available to a large community in open-source. The Polychrony environment is now integrated in the OPEES/Polarsys platform and distributed under EPL and GPL v2.0 license for the benefits of a growing community of users and contributors, among which the most active are Virginia Tech's Fermat laboratory and Inria's project-teams Aoste, Dart.

MUTANT Project-Team

4. Application Domains

4.1. Application Domains

- **Authoring and Performing Interactive Music.** The combination of both realtime machine listening systems and reactive programming paradigms has enabled the *authoring* of interactive music systems as well as their realtime performance within a coherent synchronous framework called Antescofo. The module, developed since 2008 by the team members, has gained increasing attention within the user community worldwide with more than 30 prestigious public performances yearly. The outcomes of the proposed research will enhance the interactive and reactive aspects of this emerging paradigm as well as creating novel authoring tool for such purposes. The outcome of the ANR Project INEDIT (with LABRI and GRAME and coordinated by team leader), will further extend the use-cases of *Antescofo* for interactive multimedia pieces with more complex temporal structures and computational paradigms.
- **Music Post-Production.** Outcomes of our recognition and alignment paradigms can improve and ease existing workflows employed by audio engineers for mixing and editing using commercial Digital Audio Workstations (DAW) in post-production. We have recently initiated collaborations with audio engineers at Ircam and Paris Superior Music Conservatory (CNSMDP) to define the framework [8] and we will continue to develop and integrate our tools into their daily workflow.
- **Realtime Music Information Retrieval** We will apply our information geometric approach to well-known and complex MIR problems. A glance of such problems is presented in [5]. Such applications can be used as front-end of many high-level MIR applications such as audio summarisation, audio finger printing, and automatic annotation tools. Besides such low-level enhancements, our information geometric approach can address the well-known (and still to be solved) problem of audio queries over a database.
- **Automatic Accompaniment/Creative Tools for Entertainment Industry** Technologies developed by MUTANT can find their way with general public (besides professional musicians) and within the entertainment industry. Recent trends in music industry show signs of tendencies towards more intelligent and interactive interfaces for music applications. Among them is reactive and adaptive automatic accompaniment and performance assessment as commercialized by companies such as *MakeMusic* and *Tonara*. Technologies developed around *Antescofo* can enhance interaction between user and the computer for such large public applications. We hope to pursue this by licensing our technologies to third-party companies.

PARKAS Project-Team

4. Application Domains

4.1. Application Domains

The project addresses the design, semantics and implementation of programming languages together with compilation techniques to develop provably safe and efficient computing systems. Traditional applications can be found in safety critical embedded systems with hard real-time constraints such as avionics (e.g., fly-by-wire command), railways (e.g., on board control, engine control), nuclear plants (e.g., emergency control of the plant). While embedded applications have been centralized, they are now massively parallel and physically distributed (e.g., sensor networks, train tracking, distributed simulation of factories) and they integrate computationally intensive algorithms (e.g., video processing) with a mix of hard and soft real-time constraints. Finally, systems are heterogeneous with discrete devices communicating with physical ones (e.g., interface between analog and digital circuits). Programming and simulating a whole system from a unique source code, with static guarantees on the reproducibility of simulations together with a compiler to generate target embedded code is a scientific and industrial challenge of great importance.

POP ART Project-Team

4. Application Domains

4.1. Industrial Applications

Our applications are in the embedded system area, typically: transportation, energy production, robotics, telecommunications, 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 programmation 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 our orientation towards the proposal of 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[12]) and execution platforms (OS such as VXWORKS, QNX, real-time versions of LINUX ...) starts 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[13] (based on LUSTRE), CONTROLBUILD and RT-BUILDER (based on SIGNAL) from GEENSYS[14] (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 STMicroelectronics on dynamic data-flow models of computation for streaming applications, dedicated to high definition video applications for their new STHORM manycore chip.

[12] http://www.dspaceinc.com
[13] http://www.esterel-technologies.com
[14] http://www.geensoft.com

S4 Project-Team

4. Application Domains

4.1. Modular design of embedded systems with interface theories

In 2006, with the oportunity of the SPEEDS European project on embedded system design, we decided to open a new research track on contract-/interface-based design. Our objective was to provide theory, methods and tools to support the design of embedded software in transport system industries. According to our understanding of industrial needs, gained during the SPEEDS European project, the following requirements apply to the notions of *contract* and *interface* and have been used to guide our research on this topic:

- Complex embedded and reactive systems are generally developed under a multi-layered OEM-supplier chain. Hence, a contract-based methodology should offer provision for formalizing the technical part of contractual relations. This should be achieved by formalizing, for a considered subsystem: 1/ its context of use (*assumptions*), and 2/ what is expected from the subsystem (*guarantees*). Assumptions and guarantees can be specified separately, or in a single automata-theoretic structure called interface.
- When developed under a contract-/interface-based methodology, subsystems or components should be designable in isolation, by including the needed information regarding possible future contexts of use. Subsystems or components should be substitutable to their specifications, meaning that their integration should raise no problem.
- Large systems are concurrently developed for their different *aspects* or *viewpoints* by different teams using different frameworks and tools. Examples of such aspects include the functional, reliability, timing, memory and power aspects. Each of these aspects requires specific frameworks and tools for their analysis and design. Yet, they are not totally independent but rather interact. The issue of dealing with multiple aspects or multiple viewpoints is thus essential. This implies that several contracts or interfaces are associated with a same system, sub-system, or component, namely at least one per viewpoint. These contracts/interfaces are to be interpreted in a conjunctive way and modular reasoning methods have to be developed to support large sets of contracts.
- The need for supporting conjunctive contracts/interfaces also follows from the current practice in which early requirement capture results in many elementary requirements. These requirements typically consist of English text, semi-formal languages whose sentences are translatable into predefined behavioral patterns, or even graphical scenario languages.
- It is highly desirable that designing by contracts and interfaces has the mildest possible impact on the design process, a key proprietory asset to all major companies.

4.2. Opacity, Supervision, and Petri Nets

Our activities on components emerged from a larger basis of competences developed in the past of S4 on supervisory control and Petri net synthesis. Components and their interfaces are intimately tied to supervisory control, and Petri net synthesis is a possible approach to controller synthesis. In the last four years, we have carried on work on both themes, but refocussed our research on fresh topics. A major contribution has been to study supervisory control for secrecy objectives, with promising results. Another contribution has been to study supervisory control for finite abstractions of services. The fusion of both topics, that would increase the interest of the results for Web applications, is not yet done. A different topic that we continued to investigate is the synthesis of distributed controllers based on the synthesis of distributed Petri nets. Our progress on this difficult topic is limited, but we feel we should pursue the effort.

Opacity is an abstract property that includes non-interference and that can cover confidentiality, authenticity and many other specific security concepts. Our project-team has inaugurated research on supervisory control of discrete event systems for opacity, which became soon a theme of cooperation with project-team Vertecs and subsequently attracted concurrent researches at Wayne State U., Kyoto Inst. of Tech., and U. Illinois. We have some advance over these concurrent teams.

The rest of our work on supervision focusses on minimizing communication between decentralized controllers, on asynchronous and distributed control, and on the enforcement of modal specifications. Decreasing communication between decentralized controllers was studied at Michigan U. but we could further show that minimizing communication reduces to a classical optimization problem. As regards asynchronously communicating control, the only current attempts we are aware of are those of project-teams S4 and Vertecs. As regards supervisory control w.r.t. modal specifications, the closest work is Lohmann and Wolf's synthesis of communication partners for Web services.

The approach which we propose towards distributed control relies upon the synthesis of distributed Petri nets. We have been leaders for fifteen years on the synthesis of P/T nets, on a par with the Petrify team focussed on Elementary (or safe) net synthesis. The algorithms which we have defined have been reused or adapted by many other researchers in Europe, in the US, and in China, to respond to three types of problems: controller synthesis, process mining, and concise representation of services. We are currently writing a book covering all aspects of the theory and applications of Petri net synthesis. We also pursue research on structure theory of Petri nets, in cooperation with U. Oldenburg, with focus set recently on non-interference.

4.3. Hybrid Systems Modelers

This is an opportunistic objective, not part of the plans stated when the team was formed. It results from a series of events: in 2008, Benoît Caillaud was part of the *Synchronics* large scale initiative (see section 7.1.1), dedicated to "embedded systems programming in 2020". Hybrid Systems Modelers were part of the research program. Such tools are nowadays absolutely central in the development of Cyber Physical Systems (CPS), which are physical systems in closed loop with embedded control. Hybrid Systems Modelers support the modeling of physical systems (with Ordinary Differential Equations, ODE, and Differential Algebraic Equations, DAE): Matlab-Simulink and Modelica are the main players. Our vision was that these tools should deserve similar effort in theory as synchronous languages did for the programming of embedded systems. About one year after Synchronics started (focusing mostly on other topics), the PhD thesis of Simon Bliudze came to our knowledge. This thesis contained a long chapter on the use of *non-standard analysis* as a semantic framework for hybrid systems. The exposure relied on a recent presentation of non-standard analysis, not axiomatic in style, due to the mathematician Lindström. That attracted the attention of Albert Benveniste, so he joined the group of Synchronics working on hybrid systems. This was the beginning of a deeply novel and exciting research track.

The computer science community has devoted significant efforts to the analysis and verification of hybrid automata. The framework of hybrid automata is, however, much less flexible than what actual Hybrid Systems Modelers offer. The only ongoing effort towards modeling has been developed by Edward Lee and his team as part of the Ptolemy II project. This has led to the proposal of *super-dense time semantics,* in which cascades of successive instants can occur in zero time by using $R_+ \times N$ as a time index. It turns out that the set $T = \{n\partial \mid n \in N^*\}$, where ∂ is an *infinitesimal* and N^* is the set of *non-standard integers* is such that 1/ T is dense in R_+, making it "continuous", and 2/ every $t \in T$ has a predecessor in T and a successor in T, making it "discrete" (le beurre et l'argent du beurre, as we say in french). Although non-effective from the operational point of view, the *non-standard semantics* of hybrid systems provides a framework that is very familiar to the computer scientist (who is afraid of continuous time) and at the same time efficient as a symbolic abstraction. This makes it an excellent candidate for the development of compilation schemes.

TRIO Project-Team

4. Application Domains

4.1. Domain

Three main application domains can be underlined.

- In-vehicle embedded systems. A lot of work developed in TRIO is oriented towards transportation systems (cars, autonomous vehicles, etc.). They mainly cover two points. The first one is the specification of what must be modeled in such a system and how to reach a good accuracy of a model; this leads to investigate topics like Architecture Description Languages and automatic generation of models. The second point concerns the verification of dependability properties and temporal properties required by these applications and, consequently, the development of new fault tolerant on-line mechanisms to include in an application or the automatic generation of a standard middleware.
- Compilation, memory management and low-power issues for real time embedded systems. It becomes mandatory to design embedded systems that respect performances and reliability constraints while minimizing the energy consumption. Hence, TRIO is involved, on the one hand, in the definition of ad-hoc memory management at compilation time and on the other hand, in joint study of memory management strategies and tasks scheduling for real time critical systems.
- Code analyses and software visualization for embedded systems. Despite important advances, it is still impossible to develop and optimize automatically all the programs with all their variety, especially when deployment constraints are considered. Software design and implementation thus remain highly ad-hoc, poorly automated activities, with a human being in the loop. TRIO is thus involved in the design of better tools for software engineering focusing on helping the human developer understand and develop the system, thanks to powerful automated program analyses and advanced visualizations techniques.

VERTECS Project-Team

4. Application Domains

4.1. Overview

The methods and tools developed by the VERTECS project-team for test generation and control synthesis of reactive systems are intended to be as generic as possible. This allows us to apply them in many application domains where the presence of software is predominant and its correctness is essential. In particular, we apply our research in the context of telecommunication systems, for embedded systems, for smart-cards application, and control-command systems.

4.2. Telecommunication systems

Our research on test generation was initially proposed for conformance testing of telecommunication protocols. In this domain, testing is a normalized process [21], and formal specification languages are widely used (SDL in particular). Our test generation techniques have already proved useful in this context, going up to industrial transfer. New standardized component-based design methodologies such as UML and OMG's MDE increase the need for formal techniques in order to ensure the compositionality of components, by verification and testing. Our techniques, by their genericity and adaptativity, have also proved useful at different levels of these methodologies, from component testing to system testing. The telecommunication industry now also tries to provide more and more services to the users. These services must be validated.

4.3. Software embedded systems

In the context of transport, software embedded systems are increasingly predominant. This is particularly important in automotive systems, where software replaces electronics for power train, chassis (e.g. engine control, steering, brakes) and cabin (e.g. wiper, windows, air conditioning) or new services to passengers are increasing (e.g. telematics, entertainment). Car manufacturers have to integrate software components provided by many different suppliers, according to specifications. One of the problems is that testing is done late in the life cycle, when the complete system is available. Faced with these problems, but also with the complexity of systems, compositionality of components, distribution, etc, car manufacturers now try to promote standardized interfaces and component-based design methodologies. They also develop virtual platforms which allow for testing components before the system is complete. It is clear that software quality and trust are one of the problems that have to be tackled in this context. This is why we believe that our techniques (testing and control) can be useful.

4.4. Control-command systems

The main application domain for our techniques is control-command systems. In general, such systems control costly machines (see, e.g., robotic systems, flexible manufacturing systems), that are connected to an environment (e.g., a human operator). Such systems are often critical systems and errors occurring during their execution may have dramatic economical or human consequences. In this field, the controller synthesis methodology (CSM) is useful to ensure by construction the interaction between 1) the different components, and 2) the environment and the system itself. For the first point, the CSM is often used as a safe scheduler, whereas for the second one, the supervisor can be interpreted as a safe discrete tele-operation system. Also in the context of the Vacsim ANR project, we investigate the testing, monitoring and verification of control-command systems.

ABSTRACTION Project-Team

4. Application Domains

4.1. Certification of Safety Critical Software

Absence of runtime error, Abstract interpretation, Certified compilation, Static analysis, Translation validation, Verifier.

Safety critical software may incur great damage in case of failure, such as human casualties or huge financial losses. These include many kinds of embedded software, such as fly-by-wire programs in aircrafts and other avionic applications, control systems for nuclear power plants, or navigation systems of satellite launchers. For instance, the failure of the first launch of Ariane 5 (flight Ariane 501) was due to overflows in arithmetic computations. This failure caused the loss of several satellites, worth up to $ 500 millions.

This development of safe and secure critical software requires formal methods so as to ensure that they do not go wrong, and will behave as specified. In particular, testing, bug finding methods, checking of models but not programs do not provide any guarantee that no failure will occur, even of a given type such as runtime errors; therefore, their scope is limited for certification purposes. For instance, testing can usually not be performed for *all* possible inputs due to feasibility and cost reasons, so that it does not prove anything about a large number of possible executions.

By contrast, program analysis methods such as abstract-interpretation-based static analysis are not subject to unsoundness, since they can *formally prove* the absence of bugs directly on the program, not on a model that might be erroneous. Yet, these techniques are generally incomplete since the absence of runtime errors is undecidable. Therefore, in practice, they are prone to false alarms (*i.e.*, they may fail to prove the absence of runtime errors for a program which is safe). The objective of certification is to ultimately eliminate all false alarms.

It should be noted that, due to the size of the critical codes (typically from 100 to 1000 kLOCs), only scalable methods can succeed (in particular, software model checking techniques are subject to state explosion issues). As a consequence, this domain requires efficient static analyses, where costly abstractions should be used only parsimoniously.

Furthermore, many families of critical software have similar features, such as the reliance on floating-point intensive computations for the implementation of control laws, including linear and non-linear control with feedback, interpolations, and other DSP algorithms. Since we stated that a proof of absence of runtime errors is required, very precise analyses are required, which should be able to yield no false alarm on wide families of critical applications. To achieve that goal, significant advantages can be found in the design of domain specific analyzers, such as ASTRÉE [31], [46], which has been initially designed specifically for synchronous embedded software.

Last, some specific critical software qualification procedures may require additional properties being proved. As an example, the DO-178 regulations (which apply to avionics software) require a tight, documented, and certified relation to be established between each development stage. In particular, compilation of high level programs into executable binaries should also be certified correct.

The ABSTRACTION project-team has been working on both proof of absence of runtime errors and certified compilation over the decade, using abstract interpretation techniques. Successful results have been achieved on industrial applications using the ASTRÉE analyzer. Following this success, ASTRÉE has been licensed to AbsInt Angewandte Informatik GmbH to be industrialized, and the ABSTRACTION project-team has strong plans to continue research on this topic.

4.2. Abstraction of Biological Cell Signaling Networks

Biology, Health, Static analysis.

Protein-protein interactions consist in complexations and post translational modifications such as phosphorilation. These interactions enable biological organisms to receive, propagate, and integrate signals that are expressed as proteins concentrations in order to make decisions (on the choice between cell division and cell death for instance). Models of such interaction networks suffer from a combinatorial blow up in the number of species (number of non-isomorphic ways in which some proteins can be connected to each others). This large number of species makes the design and the analysis of these models a highly difficult task. Moreover the properties of interest are usually quantitative observations on stochastic or differential trajectories, which are difficult to compute or abstract.

Contextual graph-rewriting systems allow a concise description of these networks, which leads to a scalable method for modeling them. Then abstract interpretation allows the abstraction of these systems properties. First qualitative abstractions (such as over approximation of complexes that can be built) provide both debugging information in the design phases (of models) and static information that are necessary in order to make other computations (such as stochastic simulations) scale up. Then qualitative invariants also drive efficient quantitative abstractions (such as the reduction of ordinary differential semantics).

The work of the ABSTRACTION project-team on biological cell signaling networks ranges from qualitative abstractions to quantitative abstractions.

ATEAMS Project-Team (section vide)

CARTE Project-Team

4. Application Domains

4.1. Computer Virology

4.1.1. The theoretical track.

It is rightful to wonder why there is only a few fundamental studies on computer viruses while it is one of the important flaws in software engineering. The lack of theoretical studies explains maybe the weakness in the anticipation of computer diseases and the difficulty to improve defenses. For these reasons, we do think that it is worth exploring fundamental aspects, and in particular self-reproducing behaviors.

4.1.2. The virus detection track.

The crucial question is how to detect viruses or self-replicating malwares. Cohen demonstrated that this question is undecidable. The anti-virus heuristics are based on two methods. The first one consists in searching for virus signatures. A signature is a regular expression, which identifies a family of viruses. There are obvious defects. For example, an unknown virus will not be detected, like ones related to a 0-day exploit. We strongly suggest to have a look at the independent audit [44] in order to understand the limits of this method. The second one consists in analysing the behavior of a program by monitoring it. Following [46], this kind of methods is not yet really implemented. Moreover, the large number of false-positive implies this is barely usable. To end this short survey, intrusion detection encompasses virus detection. However, unlike computer virology, which has a solid scientific foundation as we have seen, the IDS notion of "malwares" with respect to some security policy is not well defined. The interested reader may consult [65].

4.1.3. The virus protection track.

The aim is to define security policies in order to prevent malware propagation. For this, we need (i) to define what is a computer in different programming languages and setting, (ii) to take into consideration resources like time and space. We think that formal methods like rewriting, type theory, logic, or formal languages, should help to define the notion of a *formal immune system*, which defines a certified protection.

4.1.4. The experimentation track.

This study on computer virology leads us to propose and construct a "high security lab" in which experiments can be done in respect with the French law. This project of "high security lab" in one of the main project of the CPER 2007-2013.

4.2. Computations and Dynamical Systems

4.2.1. Continuous computation theories

Understanding computation theories for continuous systems leads to studying hardness of verification and control of these systems. This has been used to discuss problems in fields as diverse as verification (see e.g. [30]), control theory (see e.g. [37]), neural networks (see e.g. [66]), and so on.

We are interested in the formal decidability of properties of dynamical systems, such as reachability [56], the Skolem-Pisot problem [33], the computability of the ω-limit set [55]. Those problems are analogous to verification of safety properties.

Contrary to computability theory, complexity theory over continuous spaces is underdeveloped and not well understood. A central issue is the choice of the representation of objects by discrete data and its effects on the induced complexity notions. As for computability, it is well known that a representation is gauged by the topology it induces. However more structure is needed to capture the complexity notions: topologically equivalent representations may induce different classes of polynomial-time computable objects. For example, developing a sound complexity theory over continuous structures would enable us to make abstract computability results more applicable by analysing the corresponding complexity issues. We think that the preliminary step towards such a theory is the development of higher-order complexity, which we are currently carrying out.

In contrast with the discrete setting, it is of utmost importance to compare the various models of computation over the reals, as well as their associated complexity theories. In particular, we focus on the General Purpose Analog Computer of Claude Shannon [68], on recursive analysis [74], on the algebraic approach [63] and on computability in a probabilistic context [59].

A crucial point for future investigations is to fill the gap between continuous and discrete computational models. This is one deep motivation of our work on computation theories for continuous systems.

4.2.2. Analysis and verification of adversary systems

The other research direction on dynamical systems we are interested in is the study of properties of adversary systems or programs, i.e. of systems whose behavior is unknown or indistinct, or which do not have classical expected properties. We would like to offer proof and verification tools, to guarantee the correctness of such systems.

On one hand, we are interested in continuous and hybrid systems. In a mathematical sense, a hybrid system can be seen as a dynamical system, whose transition function does not satisfy the classical regularity hypotheses, like continuity, or continuity of its derivative. The properties to be verified are often expressed as reachability properties. For example, a safety property is often equivalent to (non-)reachability of a subset of unsure states from an initial configuration, or to stability (with its numerous variants like asymptotic stability, local stability, mortality, etc ...). Thus we will essentially focus on verification of these properties in various classes of dynamical systems.

We are also interested in rewriting techniques, as a mean to describe dynamical systems, in particular in the adversary context. As they were initially developed in the context of automated deduction, rewriting proof techniques, although now numerous, are not yet adapted to the complex framework of modelling and programming. An important stake in the domain is to enrich them to provide realistic validation tools, both in providing finer rewriting formalisms and their associated proof techniques, and in developing new validation concepts in the adversary case, i.e. when usual properties of the systems like, for example, termination are not verified.

For several years, we have been developing specific procedures for property proofs of rewriting, for the sake of programming, in particular with an inductive technique, already applied with success to termination under strategies [47], [48], [49], to weak termination [50], sufficient completeness [52] and probabilistic termination [51].

The last three results are in the context of adversary computations, since they allow to prove that a program can give the expected results, even when it diverges i.e., even when it has not the usual termination property. A common mechanism has been extracted from the above works, and provides a generic inductive proof framework for properties of reduction relations, which can be parametrized by the property to be proved [53], [54]. Provided that program code can be translated into rule-based specifications, this approach can be applied to correctness proof of software in a larger context.

A crucial element of safety and security of software systems is the problem of resources. We are working in the field of Implicit Computational Complexity. Interpretation based methods like Quasi-interpretations (QI) or sup-interpretations, are the approach we have been developing these last five years, see [60], [61], [62].

Implicit complexity is an approach to the analysis of the resources that are used by a program. Its tools come essentially from proof theory. The aim is to compile a program while certifying its complexity.

CASSIS Project-Team

4. Application Domains

4.1. Verification of Security Protocols

Security protocols such as SET, TLS and Kerberos, are designed for establishing the confidence of electronic transactions. They rely on cryptographic primitives, the purpose of which is to ensure integrity of data, authentication or anonymity of participants, confidentiality of transactions, etc.

Experience has shown that the design of those protocols is often erroneous, even when assuming that cryptographic primitives are perfect, i.e., that an encoded message cannot be decrypted without the appropriate key. An intruder can intercept, analyze and modify the exchanged messages with very few computations and therefore, for example, generate important economic damage.

Analyzing cryptographic protocols is complex because the set of configurations to consider is very large, and can even be *infinite*: one has to consider any number of sessions, any size of messages, sessions interleaving, some algebraic properties of encryption or data structures.

Our objective is to automatize as much as possible the analysis of protocols starting from their specification. This consists in designing a tool that is easy to use, enables the specification of a large number of protocols thanks to a standard high-level language, and can either look for flaws in a given protocol or check whether it satisfies a given property. Such a tool is essential for verifying existing protocols, but also for helping in designing new ones. For our tool to be easy to use, it has to provide a graphical interface allowing a user to easily perform push-button verification.

Our tools for verifying security protocols are available as components of the AVISPA platform. As an extension of the AVISPA specification language, we are working on a new environment called *CASRUL* for handling more general protocols like e-business protocols for example.

4.2. Automated Boundary Testing from Formal Specifications

We have introduced a new approach for test generation from set-oriented formal specifications: the BZ-TT method. This method is based on Constraint Logic Programming (CLP) techniques. The goal is to test every operation of the system at every boundary state using all input boundary values of that operation. It has been validated in several industry case studies for smart card OS and application validation (GSM 11-11 standard [71] and Java Card Virtual Machine Transaction mechanism [73]), information system and for embedded software [80].

This test generation method can be summed up as follows: from the formal model, the system computes boundary values to create boundary states; test cases are generated by traversal of the state space with a preamble part (sequences of operations from the initial state to a boundary state), a body part (critical invocations), an identification part (observation and oracle state computation) and a post-amble part (return path to initial or boundary state). Then, an executable test script file is generated using a test pattern and a table of correspondence between abstract operations (from the model) and concrete ones. This approach differs in several main points from previous works (e.g. [78]): first, using boundary goals as test objectives avoids the complete construction of the reachability graph; second, this process is fully automated and the test engineer could just drive it at the boundary value computation level or for the path computation.

The BZ-TT method is fully supported by the BZ-Testing-Tools tool-set. This environment is a set of tools dedicated to animation and test cases generation from B, Z or State-Chart formal specifications. It is based on the CLPS constraint solver, able to simulate the execution of the specification. By execution, we mean that the solver computes a so-called constrained state by applying the pre- and post-condition of operations. A constrained state is a constraint store where state variables and also input and output variables support constraints.

One orientation of the current work is to go beyond the finiteness assumption limitations by using symbolic constraint propagation during the test generation process. Second orientation is to extend the result to object oriented specifications as UML/OCL.Third orientation is to extend the coverage of method for security aspect.

4.3. Program Debugging and Verification

Catching bugs in programs is difficult and time-consuming. The effort of debugging and proving correct even small units of code can surpass the effort of programming. Bugs inserted while "programming in the small" can have dramatic consequences for the consistency of a whole software system as shown, e.g., by viruses which can spread by exploiting buffer overflows, a bug which typically arises while coding a small portion of code. To detect this kind of errors, many verification techniques have been put forward such as static analysis and software model checking.

Recently, in the program verification community, there seems to be a growing demand for more declarative approaches in order to make the results of the analysis readily available to the end user. To meet this requirement, a growing number of program verification tools integrate some form of theorem proving.

The goals of our research are twofold. First, we perform theoretical investigations of various combinations of propositional and first-order satisfiability checking in order to automate the theorem proving activity required to solve a large class of program analysis problems which can be encoded as first-order formulae. Second, we experimentally investigate how our techniques behave on real problems in order to make program analysis more precise and scalable. Building tools capable of providing a good balance between precision and scalability is one of the crucial challenges to transfer theorem proving technology to the industrial domains.

4.4. Verification of Web Services

Driven by rapidly changing requirements and business needs, IT systems and applications are undergoing a paradigm shift: components are replaced by services, distributed over the network, and composed and reconfigured dynamically in a demand-driven way into service-oriented architectures.Exposing services in future network infrastructures means a wide range of trust and security issues need to be adressed. Solving them is extremely hard since making the service components trustworthy is not sufficient: composing services leads to new subtle and dangerous vulnerabilities due to interference between component services and policies, the shared communication layer, and application functionality. Thus, one needs validation of both the service components and their composition into secure service architectures. In this context, there is an obvious need of applying formal methods. Our project aims at applying our proof and constraint solving techniques to reason on web services. More precisely, we focus on the composition problem in the presence of security policies.

4.5. Model-Checking of Collaborative Systems

Collaborative systems consitute a class of distributed systems where real human interactions are predominant. In these systems, users at geographically distributed sites interact by simultaneously manipulating shared objects like, text documents, XML trees, filesystems, etc. To improve data availablity, the shared objects are replicated so that the users update their local replicas and exchange their updates between them. One of the main challenges here is how to ensure the data consistency when the updates are executed in arbitrary orders at different replicas. Operational Transformation (OT) is an optimistic technique which has been proposed to overcome the consistency problem. This technique consists of an application-dependent protocol to enforce the out-of-order execution of updates even though these updates do not naturally commute. The data consistency relies crucially on the correctness of OT protocols whose proof is extremely hard. Indeed, possibly infinitely many cases should be tested. Our research work aims at applying symbolic model-checking techniques to automatically verify OT protocols. Most importantly, we are interested in finding under which conditions the model-checking problem can be reduced to a finite-state model.

CELTIQUE Project-Team (section vide)

COMETE Project-Team

4. Application Domains

4.1. Security and privacy

Participants: Miguel Andrés, Nicolás Bordenabe, Konstantinos Chatzikokolakis, Jérémy Dubreil, Catuscia Palamidessi.

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.

CONTRAINTES Project-Team

4. Application Domains

4.1. Combinatorial optimization

The number and economic impact of combinatorial optimization problems found in the industrial world are constantly increasing. They cover:

- resource allocation;
- placement, bin packing;
- scheduling;
- planning;
- transport;
- etc.

The last fifty years have brought many improvements in Operations Research resolution techniques. In this context, Constraint Programming can be seen as providing, on the one hand, constraint propagation algorithms that can be applied to various numerical or symbolic constraints, and on the other hand, declarative languages to model real-life problems and express complex resolution strategies. The latter point is crucial for designing new algorithms that cannot be defined without a sufficiently high-level language to express them. It allowed for better results than traditional methods, for instance in scheduling, and is promised to an even better future when thinking about the cooperation of global resolution, local consistency techniques and search methods.

The European FP6 Strep project Net-WMS that we have coordinated, has shown the benefit of combining discrete geometry constraints with rules to express physical, common sense and packing business constraints to solve packing problems in the context of warehouse management systems for the automotive industry. In this context, we have developed a rule-based modeling language, called Rules2CP, to express requirements in a declarative and flexible manner, and compile them to efficient constraint programs using reified constraints and a global constraint dedicated to geometrical placement problems in high dimension.

4.2. Computational Systems Biology

In partnership with biologists, we develop and experiment our modeling methods in five main leading applications:

- **Cancer chronotherapy optimization.** This research initiated in 2004 in partnership with Jean Clairambault, EPI BANG, and Francis Lévi INSERM, Hopital Paul Brousse, Villejuif, aims at understanding fundamental mechanisms involved in cancer and chronotherapies through mathematical modeling. Following the EU STREP project (2006-2009) on “temporal genomics for patient tailored chronotherapeutics”, coordinated by Francis Lévi, and in the framework of the Era-Net SysBio C5Sys project (2010-2013) coordinated by Francis Lévi and David Rand, University of Warwick, UK, we develop coupled models of the cell cycle, the circadian clock, the DNA repair system, irinotecan metabolism and drug injection optimization, focussing on the interactions between the cell cycle and the circadian clock in mammalian cells.
- **Mammalian cell cycle regulation.** This theme that is closely related to the previous one has lead to a formal collaboration in the framework of the ANR Syscomm project CALAMAR, started in 2009 on the “Compositional modeling and Analysis of LArge MoleculAr Regulatory networks”. In partnership with Claudine Chaouiya, TAGC INSERM, Marseille, and Laurence Calzone, Institut Curie, Paris, this project aims at applying our computational techniques – both qualitative and quantitative – to the analysis of the large scale RB/E2F network, in order to elucidate various features of the human cell proliferation, especially in the case of healthy and bladder-tumor cells of different aggressiveness.

- **G-protein coupled receptor signal transduction**. This research initiated in 2004 in partnership with Eric Reiter, INRA Tours, and Frédérique Clément, EPI SISYPHE, aimed at understanding the structure and the dynamics of the follicule stimulating hormone (FSH) and angiotensine signal transduction in mammalian cells. It was first conducted in the INRA AgroBi project INSIGHT (2006-2009) and in the AE REGATE.

  The article [4] concludes our fruitful collaboration over this period of eight years, with a tightly coupled formal and experimental study of GPCR signaling, of particular importance in medicine since these receptors are the most common drug target.

- **Real-time control of gene expression in yeast.** This research lead in the team by Grégory Batt investigates the possibilities to control gene expression in living cells. In collaboration with Pascal Hersen and Samuel Bottani, biophysicists at the Matière and Systèmes Complexes lab, CNRS/Paris Diderot University, we develop a microfluidic platform and control software for the real-time control of gene expression in yeast. In a larger initiative, we consider a similar problem but in mammalian cells, where the stochasticity of gene expression makes the control problem particularly challenging. The Iceberg Investissement d'Avenir project, coordinated by Grégory Batt, involves the MSC, BM2A, LIFL and PPS labs, and the Jacques Monod Institut. Similarly, the Contraintes research group is also involved in the Inria/INSERM large-scale initiative action COLAGE coordinated by Huges Berry, EPI COMBINING, with François Taddei, Ariel Lindner, INSERM Paris Necker, Hidde de Jong, Delphine Ropers, EPI IBIS, Jean-Luc Gouzé, and Madalena Chaves, EPI COMORE. In this project, we investigate the possibilities to control and reprogram growth and aging in bacteria *E. coli* using synthetic biology approaches.

- **Artificial tissue homeostasis in mammalian cells.** Artificial tissue design is a particularly challenging problem in synthetic biology since the system behavior results from the interplay between intra- and intercellular dynamics. In the framework of the Syne2arti ANR project, coordinated by Grégory Batt, and involving Dirk Draso, EPI BANG, Oded Maler, CNRS Verimag, and Ron Weiss, MIT, USA, we design and genetically-engineer mammalian cells to obtain a tissue having a desired cell density. The long-term correct functioning of the system relies several key aspects, including individual cell decisions, collective, spatial aspects, and cell-to-cell variability.

- **TGFβ signaling and initiation of translation in sea urchin.** In the framework of the BioTempo ANR project, we recently started to apply the different algorithms available in the BIOCHAM platform to the modeling of the TGFβ signaling network in collaboration with the SeRAIC lab (Rennes, France) and of the sea urchin's initiation of translation with Laboratoire Mer et Santé (Roscoff, France). In the first case, the main challenge is to compare and understand crosstalks between the SMAD-dependent fast pathway and the MAPK-dependent slower pathway that is often related to cancer. In the second case there is a whole issue of parametrization even for small models since the data is quite sparse. The different parameter learning features of BIOCHAM, notably based on temporal logics, are therefore put to good use.

DEDUCTEAM Team

3. Application Domains

3.1. Safety of Aerospace systems

In parallel with this effort in logic and in the development of proof checkers and automated theorem proving systems, we always have been interested in using such tools. One of our favorite application domain is the safety of aerospace systems. Together with César Munoz' team in Nasa-Langley, we have proved the correctness of several geometric algorithms used in air traffic control.

This has led us sometimes to develop such algorithms ourselves, and sometimes to develop tools for automating these proofs.

3.2. B-set theory

The B method allows the user to develop software correct by construction, going from abstract models to implementations via refinement. During the development process, proof obligations are generated. The formalism underlying B is based on predicate logic and B-set theory. Atelier B that supports the B method provides interactive and automatic provers. To increase automation the user may add proof rules which, if not correct, may corrupt the process. Siemens has developed a tool chain to verify such added proof rules. In particular we have to verify that any proof rule derives from B logic. This step has to be as automatic as possible. Furthermore confidence in these verification proofs is required. A first attempt using the first order prover Zenon allowed the verification of a large number of proof rules [10]. To go further we have experimented techniques such as super deduction and deduction modulo. B-set theory is an interesting benchmark for the tools developed by Deducteam since this theory contains numerous operators and predicates defined by equations or rewrite rules.

FORMES Team

4. Application Domains

4.1. Simulation

Simulation is relevant to most areas where complex embedded systems are used, not only to the semiconductor industry for System-on-Chip modeling, but also to any application where a complex hardware platform must be assembled to run the application software. It has applications for example in industry automation, digital TV, telecommunications and transportation.

4.2. Certified Compilation for Embedded systems

Many frameworks have been designed in order to make the design and the development of embedded systems more rigourous and secure on the basis of some formal model. All these frameworks implicitly assume the *reliability of the translation* to executable code, in order to guarantee the verified properties in the design level are preserved in the implementation. In other words, they rely on a claim saying that the compilers from high level model description to the implementation perfectly will not introduce undesired behaviors or errors in silence. The only safe way to satisfy such a claim is to certify correctness of the compilers, that is, to prove that the code they produce has exactly the semantics of the source code or model.

4.3. Distributed Systems

Many embedded systems run in a distributed environment. Distributed systems raise extremely challenging issues, both for the design and the implementation, because decisions can be made only from a local knowledge, which is imperfect due to communication time and unreliability of transmissions.

4.4. Security

The convergence between embedded technologies and the Internet offers many opportunities to malicious people for breaking the privacy of consumers or of organisations. Using cryptography is not enough for ensuring the protection of data, because of possible flaws in protocols and interfaces, providing opportunities for many well-known attacks. This area is therefore an important target of formal methods.

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 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 Caml 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 references, 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 Caml 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 [48] and enforcement of data confidentiality through type-based inference of information flows and noninterference properties [51].

4.3. Processing of complex structured data

Like most functional languages, Caml 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. Languages such as CDuce and OCamlDuce extend these benefits to the handling of semi-structured XML data [44]. Therefore, Caml 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 Caml 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 Caml 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 Caml language has an impact on the teaching of programming. Caml Light 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, USA, and Japan.

MARELLE Project-Team (section vide)

MEXICO Project-Team

4. Application Domains

4.1. Panorama

telecommunications, multimedia, transportation systems, web services

MExICo's research is motivated by problems on system management in several domains:

- 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.

This list is likely to grow over the next years as we continue our research.

4.2. Autonomous Telecommunications Systems: In-Band Supervision

Participants: Stefan Haar, Serge Haddad.

In the context of traditional hard-wired communication networks, supervision structures for managing faults, configuration, provisioning etc could be developed with a fixed infrastructure, and perform the communication between sensors, supervisors, policy enforcement points etc over a separate network using separate hardware. This rigid, **out-of-band** technology does not survive passing to today's and tomorrow's services and networks. In fact, the dynamic mobility of services combined across sites and domains cannot be captured unless the network used for supervision evolves in the same way and simultaneously, which rules out static solutions; but providing out-of-band infrastructure that grows with the networks to be supervised would be prohibitively expensive, if at all technically feasible. *Heterogeneity* is the other feature of modern networks that forces a change, since different domains are not likely to agree on a pervasive third-party supervision. Rather, the providers will keep control over the internal state and evolution of their domain, and accept only exchange through standardized outward interfaces.

Supervision has thus to be re-invented on an *in-band*, *autonomous* base: monitoring probes deployed on the web, dysfunctions on one peer node diagnosed by another peer in a network with changing configuration, enhanced supervisor and actor capacities of services, etc. *MExICo* will work on improving the interoperability of service components through continued application of e.g. distributed techniques for control and diagnosis.

4.3. Web Services

Participants: Stefan Haar, Serge Haddad.

Specific applications targeted by *MExICo* include the problem of adaptation in Service-Oriented Computing (SOC). The challenge is here twofold, stemming both from the distributed nature of services (scattered over the entire web) and their heterogeneous origins.

4.3.1. Context

Web services have become the most frequently used model of design and programming based on components for business applications. Web service languages like BPEL have useful constructors that manage for instance exceptions, (timed guarded) waiting of messages, parallel execution of processes, distant service invocations, etc. Interoperability of components is based on interaction protocols associated with them and often published on public or private registers. In the framework of Web services, these protocols are called abstract processes by contrast with business processes (i.e. services). Composition of components must be analyzed for several reasons and at least to avoid deadlocks during execution. This has led to numerous works that focus on compositional verification, substitution of a component by another one, synthesis of adaptators, etc., and triggered a push towards a unifying theoretical framework (see e.g. [115], [119])

4.3.2. Problems

Interoperability requires that when a user or a program wants to interact with the component, the knowledge of the interaction protocol is enough. Our previous works have shown that the interaction protocols can be inherently ambiguous: no client can conduct a correct interaction with the component in every scenario. This problem is even more complex when the protocol can evolve during execution due to adaptation requirements. The composition of components also raises interesting problems. When composing optimal components (w.r.t. the number of states for instance) the global component can be non optimal. So one aims at reducing a posteriori or better on the fly the global component. At last, the dynamical insertion of a component in a business process requires to check whether this insertion is behaviorally consistent [121], [109]

We do not intend to check global properties based on a modular verification technique. Rather, given an interaction protocol per component and a global property to ensure, we want to synthesize an adaptator per component such that this property is fulfilled or to detect that there cannot exist such adaptators [106]. In another research direction, one can introduce the concept of utility of a service and then optimize a system i.e. keeping the same utility value while reducing the resources (states, transitions, clocks, etc.).

PAREO Project-Team

4. Application Domains

4.1. Application Domains

Beside the theoretical transfer that can be performed via the cooperations or the scientific publications, an important part of the research done in the *Pareo* group team is published within software. *Tom* is our flagship implementation. It is available via the Inria Gforge (http://gforge.inria.fr) and is one of the most visited and downloaded projects. The integration of high-level constructs in a widely used programming language such as Java may have an impact in the following areas:

- Teaching: when (for good or bad reasons) functional programming is not taught nor used, *Tom* is an interesting alternative to exemplify the notions of abstract data type and pattern-matching in a Java object oriented course.
- Software quality: it is now well established that functional languages such as Caml are very successful to produce high-assurance software as well as tools used for software certification. In the same vein, *Tom* is very well suited to develop, in Java, tools such as provers, model checkers, or static analyzers.
- Symbolic transformation: the use of formal anchors makes possible the transformation of low-level data structures such as C structures or arrays, using a high-level formalism, namely pattern matching, including associative matching. *Tom* is therefore a natural choice each time a symbolic transformation has to be implemented in C or Java for instance. *Tom* has been successfully used to implement the Rodin simplifier, for the B formal method.
- Prototyping: by providing abstract data types, private types, pattern matching, rules and strategies, *Tom* allows the development of quite complex prototypes in a short time. When using Java as the host-language, the full runtime library can be used. Combined with the constructs provided by *Tom*, such as strategies, this procures a tremendous advantage.

One of the most successful transfer is certainly the use of *Tom* made by Business Objects/SAP. Indeed, after benchmarking several other rule based languages, they decided to choose *Tom* to implement a part of their software. *Tom* is used in Paris, Toulouse and Vancouver. The standard representation provided by *Tom* is used as an exchange format by the teams of these sites.

PARSIFAL Project-Team

4. Application Domains

4.1. Automated Theorem Proving

Automated theorem proving has traditionally focused on classical first-order logic, but non-classical logics are increasingly becoming important in the specification and analysis of software. Most type systems are based on (possibly second-order) propositional intuitionistic logic, for example, while resource-sensitive and concurrent systems are most naturally expressed in linear logic.

The members of the Parsifal team have a strong expertise in the design and implementation of performant automated reasoning systems for such non-classical logics. In particular, the Linprover suite of provers [35] continue to be the fastest automated theorem provers for propositional and first-order linear logic.

Any non-trivial specification, of course, will involve theorems that are simply too complicated to prove automatically. It is therefore important to design semi-automated systems that allow the user to give high level guidance, while at the same time not having to write every detail of the formal proofs. High level proof languages in fact serve a dual function – they are more readily comprehended by human readers, and they tend to be more robust with respect to maintenance and continued evolution of the systems. Members of the Parsifal team, in association with other Inria teams and Microsoft Research, have been building a heterogeneous semi-automatic proof system for verifying distributed algorithms [36].

On a more foundational level, the team has been developing many new insights into the structure of proofs and the proof search spaces. Two directions, in particular, present tantalizing possibilities:

- The concept of *multi-focusing* [37] can be used to expose concurrency in computational behavior, which can in turn be exploited to prune areas of the proof search space that explore irrelevant interleavings of concurrent actions.
- The use of *bounded search*, where the bounds can be shown to be complete by meta-theoretic analysis, can be used to circumvent much of the non-determinism inherent in resource-sensitive logics such as linear logic. The lack of proofs of a certain bound can then be used to justify the presence or absence of properties of the encoded computations.

Much of the theoretical work on automated reasoning has been motivated by examples and implementations, and the Parsifal team intends to continue to devote significant effort in these directions.

4.2. Mechanized Metatheory

There has been increasing interest in the use of formal methods to provide proofs of properties of programs and programming languages. Tony Hoare's Grand Challenge titled "Verified Software: Theories, Tools, Experiments" has as a goal the construction of "verifying compilers" for a world where programs would only be produced with machine-verified guarantees of adherence to specified behavior. Guarantees could be given in a number of ways: proof certificates being one possibility.

The POPLMark challenge [33] envisions "a world in which mechanically verified software is commonplace: a world in which theorem proving technology is used routinely by both software developers and programming language researchers alike." The proposers of this challenge go on to say that a "crucial step towards achieving these goals is mechanized reasoning about language metatheory."

The Parsifal team has developed several tools and techniques for reasoning about the meta-theory of programming languages. One of the most important requirements for programming languages is the ability to reason about data structures with binding constructs up to α-equivalence. The use of higher-order syntax and nominal techniques for such data structures was pioneered by Miller, Nadathur and Tiu. The Abella system (see Section 3.2) implements a refinement of a number of these ideas and has been used to give full solutions to sections of the POPLMark challenge in addition to fully formal proofs of a number of other theorems in the meta-theory of the λ-calculus.

Now that the Abella system has been in circulation among colleagues during the past couple of years, there are many aspects of the methodology that now need to be addressed. During the summer of 2011, the team employed three interns Carnegie Mellon University and McGill University to work on different aspects of Abella. Particular focus was given to better ways to manipulate specification-logic contexts in the reasoning-logic and with finding ways to have Abella output a proper proof object (different from the scripts that are used to find a proof).

Our colleague Alwen Tiu from the Australian National University has also been building on our Bedwyr model checking tool so that we can build on top of it his SPEC system for doing model checking of spi-calculus expressions. We have adopted his enhancements to Bedwyr and are developing further improvements within the context of the BATT project (see Section 5.2).

4.3. Proof Certificates

Members of the Parsifal team have shown how to specify a large variety of proof systems—including natural deduction, the sequent calculus, and various tableau and free deduction systems—uniformly using either focused linear logic [52], [50] or focused intuitionistic logic [43] as the meta-language. In the presence of induction and co-induction, arbitrary finite computations can be embedded into single synthetic steps [34]. Additional work [8] shows that this same framework can also capture resolution refutations as well as Pratt primality certificates.

An important application then of this work in designing synthetic inference systems based on classical and intuitionistic logic is that of designing a *broad spectrum proof certificate*. The definition of proof certificates can be remarkably flexible within the simple setting of focused proofs.

The most important implications of such a certificate format would be that most of the worlds theorem provers should be able to print out their proofs and communication them to other provers: these other provers could then check such certificates by expanding the synthetic connectives they contain down into a small and fixed set of “micro” inference rules.

4.4. Automated reasoning and SMT solving

Automated reasoning uses a broad range of techniques whose soundness and completeness relate to the existence of proofs. The research programme of the ANR PSI project at Parsifal is to build a finer-grained connection by specifying automated reasoning techniques as the step-by-step construction of proofs, as we know it from proof theory and logic programming. The goal is to do this in a unifying framework, namely proof-search in a polarized and focussed logic. One of the advantages of this is to combine those techniques more easily. Another one is to envisage extending those techniques.

For instance, SAT-modulo-Theory problems require the combination of logical reasoning with domain-specific decision procedures. So in the PSI project we study how to incorporate the call to decision procedures in proof-theoretical framework like the focussed sequent calculus, and the proof-search mechanisms that are related to it.

In the same spirit we also study how to handle existential variables and equality, for which specific automated reasoning techniques have been designed (superposition / paramodulation calculi).

PI.R2 Project-Team

4. Application Domains

4.1. The impact of Coq

Coq is one of the 8 most used proof assistants in the world. In Europe, its main challengers are Isabelle (developed in Munich, Germany), HOL (developed in Cambridge, UK) and Mizar (developed in Białystok, Poland).

Coq is used in various research contexts and in a few industrial contexts. It is used in the context of formal mathematics at the University of Nijmegen (constructive algebra and analysis), Inria Sophia-Antipolis (number theory and algebra), Inria-MSR joint lab (group theory), the University of Nice (algebra). It is used in France in the context of computer science at Inria-Rocquencourt (certified compilation), Inria-Saclay (certification of imperative programs), LORIA, Strasbourg (certification of geometry algorithms). Outside France, it is used in the context of computer science e.g. at U. Penn, Harvard (programming languages, semantics), Yale, Ottawa and Berkeley Universities (building of a certified platform for proof-carrying code), University of Princeton (certified compilation), AIST at Tokyo (certification of cryptographic protocols), Microsoft Research Cambridge (proof of imperative programs), ... In the industry, it is used by Gemalto and Trusted Logic (JavaCard formal model and commercial applets certification).

All in all, it is difficult to evaluate how much Coq is used. Two indicators are the readership of the textbook on Coq by Yves Bertot and Pierre Castéran [35] and the number of subscribers to the Coq-club mailing list. More than 1200 copies of the book have been sold. There has been a second printing , and a Chinese translation of the book has been published. There are around 600 subscribers to the mailing list. Coq is taught or used for teaching in many universities: Paris, Bordeaux, Lyon, Nice, Strasbourg, CNAM, Nottingham, Ottawa, U. Penn, Harvard, MIT, Princeton, Yale, Berkeley, Warsaw, Krakow, Rosario in Argentina, ...

Users of the assistant are also disseminating the use of the tool: A collaborative effort led by B. Pierce's team at U. Penn gave rise to a set of courses named Software Foundations (http://www.seas.upenn.edu/~cis500/current/sf/index.html) [66] on basic logic and computer science in Coq, that is used by many universities and individuals throughout the world. A. Chlipala wrote an advanced textbook on "Certified Programming with Dependent Types" in Coq, freely available on the web and soon to be published by MIT Press [37].

PROSECCO Project-Team

4. Application Domains

4.1. Cryptographic protocol implementations

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, 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 authenticate authorize users using a single sign-on protocol such as OAuth, a cloud storage service may encrypt user files on the server-side using XML encryption, and a password manager may encrypt passwords in the browser using a JavaScript cryptographic library. We build verification tools that can analyze such usages in commercial web applicaitons and evaluate their security against sophisticated web-based attacks.

SECSI Project-Team

4. Application Domains

4.1. Application Domains

Here are a few examples of applications of research done in SECSI:

- Security of electronic voting schemes: the case of the Helios protocol, used in particular at University of Louvain-la-Neuve (2010) and at the International Association for Cryptographic Research (IACR).
- Security of the protocols involved in the TPM (Trusted Platform Module) chip, a chip present in most PC laptops today, and which is meant to act as a trusted base.
- Security of the European electronic passport—and the discovery of an attack on the French implementation of it.
- The Tookan tool allows one to assess the security of security tokens. These tokens are meant as safes holding secret keys, which should never be permitted to get out unencrypted. Several vulnerabilities discovered. Several interesting customers in banking (Barclays), in aeronautics (Boeing), notably.
- Intrusion detection with the Orchids tool: several interested partners, among which EADS Cassidian, Thales, Galois Inc. (USA), the French Direction Générale de l'Armement (DGA).

TASC Project-Team

4. Application Domains

4.1. Introduction

Constraint programming deals with the resolution of decision problems by means of rational, logical and computational techniques. Above all, constraint programming is founded on a clear distinction between, on the one hand the description of the constraints intervening in a problem, and on the other hand the techniques used for the resolution. The ability of constraint programming to handle in a flexible way heterogeneous constraints has raised the commercial interest for this paradigm in the early nighties. Among his fields of predilection, one finds traditional applications such as computer aided decision-making, scheduling, planning, placement, logistics or finance, as well as applications such as electronic circuits design (simulation, checking and test), DNA sequencing and phylogeny in biology, configuration of manufacturing products or web sites, formal verification of code.

4.2. Panorama

In 2012 the TASC team was involved in the following application domains:

- *Planning and replanning* in Data Centres (SelfXL project).
- *Packing complex shapes* in the context of a warehouse (NetWMS2 project).
- Building decision support system for *city development planning with evaluation of energy impacts* (SUSTAINS project).
- *Optimizing electricity production* in the context of the Gaspard Monge call program for Optimisation and Operation Research. We extract global constraints from daily energy production temporal series issued from all productions plants of EDF over a period of several years.

TOCCATA Team

4. Application Domains

4.1. Application Domains

Keywords: embedded software, smartcards, avionics, telecommunication, transportation systems

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. The domains of application include

- Transportation: aeronautics, railroad, space flight, automotive
- Communications: mobile phones, smart phones, Web applications
- Financial applications, banking
- Medicine: diagnostic devices, computer-assisted surgery
- Databases with confidentiality requirements (e.g. health records, electronic voting)

Currently our industrial collaborations mainly belong the first of these domains: transportation. These include, in the context of the ANR U3CAT project (Airbus France,Toulouse; Dassault Aviation, Saint-Cloud; Sagem Défense et Sécurité):

- proof of C programs via *Frama-C*/*Jessie*/*Why* ;
- proof of floating-point programs ;
- use of the *Alt-Ergo* prover via CAVEAT tool (CEA) or *Frama-C*/WP.

In the context of the FUI project Hi-Lite, the Adacore (Paris) uses *Why3* and *Alt-Ergo* as back-endd to GnatProve, an environment for verification of Ada programs. This is applied in the domain of aerospace (Thales).

In the context of a new ANR project BWare, we investigate the use of *Why3* and *Alt-Ergo* as an alternative back-end for checking proof obligation generated by *Atelier B*, whose main applications are railroad-related software (http://www.methode-b.com/documentation_b/ClearSy-Industrial_Use_of_B.pdf, collaboration with Mitsubishi Electric R&D Centre Europe, Rennes; ClearSy, Aix-en-Provence)

Apart from the domain of transportation, the Cubicle model checker modulo theoriesbased on the *Alt-Ergo* SMT prover (collaboration with Intel Strategic Cad Labs, Hillsboro, OR, USA) can be applied to verification of concurrent programs and protocols (http://cubicle.lri.fr/).

TYPICAL Project-Team (section vide)

VERIDIS Project-Team

4. Application Domains

4.1. Application Domains

Our work focuses on distributed algorithms and protocols. These are or will be found at all levels of computing infrastructure, from many-core processors and systems-on-chip to wide-area networks. We are particularly interested in novel paradigms, for example ad-hoc networks that underly mobile and low-power computing or overlay networks and peer-to-peer networking that provide services for telecommunication or cloud computing services. Distributed protocols underly computing infrastructure that must be highly available and mostly invisible to the end user, therefore correctness is important. 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 work together with domain experts on the modeling and verification of these protocols. These cooperations help us focus on concrete algorithms and ensure that our work is relevant to the distributed algorithm community.

Formal verification techniques that we study can contribute to certify 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 require code analysis, based on mathematically precise foundations. While initially the requirements of certified development have mostly been restricted to safety-critical systems, they are becoming more and more common due to the cost associated with malfunctioning system components and software. We are in particular working on modeling and verifying medical devices that require closed-loop models of both the system and its environment.