RESEARCH CENTER
Paris - Rocquencourt

FIELD

Activity Report 2012

Section Application Domains

Edition: 2013-04-24

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.

ALPAGE Project-Team

4. Application Domains

4.1. Panorama

NLP tools and methods have many possible domains of application. Some of then are already mature enough to be commercialized. They can be roughly classified in three groups:

Human-computer interaction : mostly speech processing and text-to-speech, often in a dialogue context; today, commercial offers are limited to restricted domains (train tickets reservation...);

Language writing aid : spelling, grammatical and stylistic correctors for text editors, controlled-language writing aids (e.g., for technical documents), memory-based translation aid, foreign language learning tools, as well as vocal dictation;

Access to information : tools to enable a better access to information present in huge collections of texts (e.g., the Internet): automatic document classification, automatic document structuring, automatic summarizing, information acquisition and extraction, text mining, question-answering systems, as well as surface machine translation. Information access to speech archives through transcriptions is also an emerging field.

Experimental linguistics : tools to explore language in an objective way (this is related, but not limited to corpus linguistics).

Alpage focuses on some applications included in the three last points, such as information extraction and (linguistic and extra-linguistic) knowledge acquisition (4.2), text mining (4.3), spelling correction (4.5) and experimental linguistics (4.6).

4.2. Information extraction and knowledge acquisition

Participants: Éric Villemonte de La Clergerie, Rosa Stern, François-Régis Chaumartin, Benoît Sagot.

The first domain of application for Alpage parsing systems is information extraction, and in particular knowledge acquisition, be it linguistic or not, and text mining.

Knowledge acquisition for a given restricted domain is something that has already been studied by some Alpage members for several years (ACI Biotim, biographic information extraction from the Maitron corpus, SCRIBO project). Obviously, the progressive extension of Alpage parsing systems or even shallow processing chains to the semantic level increase the quality of the extracted information, as well as the scope of information that can be extracted. Such knowledge acquisition efforts bring solutions to current problems related to information access and take place into the emerging notion of *Semantic Web*. The transition from a web based on data (textual documents,...) to a web based on knowledge requires linguistic processing tools which are able to provide fine grained pieces of information, in particular by relying on high-quality deep parsing. For a given domain of knowledge (say, news or tourism), the extraction of a domain ontology that represents its key concepts and the relations between them is a crucial task, which has a lot in common with the extraction of linguistic information.

In the last years, such efforts have been targeted towards information extraction from news wires in collaboration with the Agence France-Presse (Rosa Stern is a CIFRE PhD student at Alpage and at AFP, and works in relation with the ANR project EDyLex) as well as in the context of the collaboration between Alpage and Proxem, a startup created by François-Régis Chaumartin, PhD student at Alpage (who has defended his PhD in 2012).

These applications in the domain of information extraction raise exciting challenges that require altogether ideas and tools coming from the domains of computational linguistics, machine learning and knowledge representation.

4.3. Processing answers to open-ended questions in surveys: vera

Participants: Benoît Sagot, Valérie Hanoka.

Verbatim Analysis is a startup co-created by Benoît Sagot from Alpage and Dimitri Tcherniak from Towers Watson, a world-wide leader in the domain of employee research (opinion mining among the employees of a company or organization). The aim of its first product, *vera*, is to provide an all-in-one environment for editing (i.e., normalizing the spelling and typography), understanding and classifying answers to open-ended questions, and relating them with closed-ended questions, so as to extract as much valuable information as possible from both types of questions. The editing part relies in part on SXPipe (see section 5.6) and Alexina morphological lexicons. Several other parts of *vera* are co-owned by Verbatim Analysis and by Inria.

4.4. Multilingual terminologies and lexical resources for companies

Participants: Éric Villemonte de La Clergerie, Mickael Morardo, Benoît Sagot.

Lingua et Machina is a small company now headed by François Brown de Colstoun, a former Inria researcher, that provides services for developing specialized multilingual terminologies for its clients. It develops the WEB framework Libellex for validating such terminologies. A formal collaboration with ALPAGE has been set up, with the recruitment of Mikael Morardo as engineer, funded by Inria's DTI. He works on the extension of the web platform *Libellex* for the visualization and validation of new types of lexical resources. In particular, he has integrated a new interface for handling monolingual terminologies, lexical networks, and bilingual wordnet-like structures.

4.5. Automatic and semi-automatic spelling correction in an industrial setting

Participants: Benoît Sagot, Éric Villemonte de La Clergerie, Laurence Danlos.

NLP tools and resources used for spelling correction, such as large n-gram collections, POS taggers and finite-state machinery are now mature and precise. In industrial setting such as post-processing after large-scale OCR, these tools and resources should enable spelling correction tools to work on a much larger scale and with a much better precision than what can be found in different contexts with different constraints (e.g., in text editors). Moreover, such industrial contexts allow for a non-costly manual intervention, in case one is able to identify the most uncertain corrections. An FUI project on this topic has been proposed in collaboration with Diadeis, a company specialized in text digitaliNzation, and two other partners. It has been rerouted to the "Investissements d'avenir" framework, and has been accepted. It started in 2012.

4.6. Experimental linguistics

Participants: Benoît Crabbé, Juliette Thuilier, Luc Boruta.

Alpage is a team that dedicates efforts in producing ressources and algorithms for processing large amounts of textual materials. These ressources can be applied not only for purely NLP purposes but also for linguistic purposes. Indeed, the specific needs of NLP applications led to the development of electronic linguistic resources (in particular lexica, annotated corpora, and treebanks) that are sufficiently large for carrying statistical analysis on linguistic issues. In the last 10 years, pioneering work has started to use these new data sources to the study of English grammar, leading to important new results in such areas as the study of syntactic preferences [60], [128], the existence of graded grammaticality judgments [83].

The reasons for getting interested for statistical modelling of language can be traced back by looking at the recent history of grammatical works in linguistics. In the 1980s and 1990s, theoretical grammarians have been mostly concerned with improving the conceptual underpinnings of their respective subfields, in particular through the construction and refinement of formal models. In syntax, the relative consensus on a generative-transformational approach [71] gave way on the one hand to more abstract characterizations of the language faculty [71], and on the other hand to the construction of detailed, formally explicit, and often implemented, alternative formulation of the generative approach [59], [94]. For French several grammars have been implemented in this trend, among which the tree adjoining grammars of [63], [73] among others. This general movement led to much improved descriptions and understanding of the conceptual underpinnings of both linguistic competence and language use. It was in large part catalyzed by a convergence of interests of logical, linguistic and computational approaches to grammatical phenomena.

However, starting in the 1990s, a growing portion of the community started being frustrated by the paucity and unreliability of the empirical evidence underlying their research. In syntax, data was generally collected impressionistically, either as ad-hoc small samples of language use, or as ill-understood and little-controlled grammaticality judgements (Schütze 1995). This shift towards quantitative methods is also a shift towards new scientific questions and new scientific fields. Using richly annotated data and statistical modelling, we address questions that could not be addressed by previous methodology in linguistics. In this line, at Alpage we have started investigating the question of choice in French syntax with a statistical modelling methodology. Currently two studies are being led on the position of attributive adjectives w.r.t. the noun and the relative position of postverbal complement. This research has contributed to establish new links with the Laboratoire de Linguistique Formelle (LLF, Paris 7) and the Laboratoire de Psychologie et Neuropsychologie Cognitives (LPNCog, Paris 5).

On the other hand we have also started a collaboration with the Laboratoire de Sciences Cognitives de Paris (LSCP/ENS) where we explore the design of algorithms towards the statistical modelling of language acquisition (phonological acquisition). This is currently supported by one PhD project.

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.

ARLES Project-Team

4. Application Domains

4.1. Application Domains

The ARLES project-team is interested in the application of pervasive computing, and as such considers various application domains. Indeed, our application domain is voluntarily broad since we aim at offering generic solutions. However, we examine exploitation of our results for specific applications, as part of the experiments that we undertake to validate our research results through prototype implementation. Applications that we consider in particular include demonstrators developed in the context of the European and National projects to which we contribute (§ 7.1 & 7.2).

AXIS Project-Team

3. Application Domains

3.1. Panorama: Living Labs, Smart Cities

AxIS addresses applicative field which has the following features:

a) requiring usage/data storage, preprocessing and analysis tools

- for designing, evaluating and improving huge evolving hypermedia information systems (mainly Web-based ISs), for which end-users are of primary concern,
- for a better understanding of service/product used with data mining techniques and knowledge management
- for social network analysis (for example in Web 2.0 applications, Business Intelligence, Sustainable Development, etc.)

b) requiring user-driven innovation methods.

Even if our know-how, methods and algorithms have a cross domain applicability, our team chooses to focus on **Living Lab projects** (and mainly related to **Sustainable Development for Smart Cities**) (cf. section 5.5.5 which imply user involvement for the generation of future services/products. Indeed, following the Rio Conference (1992) and the Agenda for the 21st Century, local territories are now directly concerned with the set up of actions for a sustainable development. In this frame, ICT tools are supposed to be very efficient to re-engage people in the democratic process and to make decision-making more transparent, inclusive and accessible. So, sustainable development is closely associated with citizen participation. The emerging research field of e-democracy (so called Digital Democracy or eParticipation), concerned with the use of communications technologies such as the Internet to enhance the democratic processes is now a very active field. Though still in its infancy, a lot of literature is already available (see for instance: http://itc.napier.ac.uk/ITC/publications.asp or http://www.demo-net.org/ for a global view of work in Europe) and numerous different topics are addressed in the field.

Our experience particularly stressed on the following applicative domains:

- Transportation systems & Mobility (cf. section 3.2),
- Tourism (cf. section 3.3),
- User Involvement in Energy, Environment, Well Being & Health and e-governement (cf. section 3.4).

3.2. Transportation Systems & Mobility

Major recent evolutions in Intelligent Transportation Systems (ITS) are linked to rapid changes in communication technologies, such as ubiquitous computing, semantic web, contextual design. A strong emphasis is now put on mobility improvements. In addition to development of sustainable transportation systems (better ecological vehicles' performance, reduction of impacts on town planning ...) these improvements concern also mobility management, that is specific measures to encourage people to adopt new mobility behaviour such as public transportation services rather than their personal car. These prompting measures concern for instance the quality of traveller's information systems for trip planning, the ability to provide real time recommendations for changing transportation means according to traffic information, and the quality of embedded services in vehicles to provide enhanced navigation aids with contextualised and personalised information.

Since 2004, AxIS has been concerned with mobility projects :

- PREDIT (2004-2007): The MobiVIP project has been an opportunity to collaborate with local Institutions (Communauté d'Agglomération de Sophia Antipolis - CASA) and SMEs (VU Log) and to apply AxIS' know-how in data and web mining to the field of transportation systems.
- Traveller's information systems and recommender systems have been studied with the evaluation of two CASA web sites : the "Envibus" web site which provides information about a bus network and the "Otto&co" web site support car-sharing.
- Advanced transportation systems has been studied in PREDIT TIC TAC (2010-2012): this project (cf. section 6.1.1) aimed at optimizing travel time by providing in an aera with weak transportation services, a just in time on demand shuttle based on real time information. It was for AxIS the opportunity to experiment user implication in the design of a new travel information system called MOBILTIC
- User Experience: in the ELLIOT project (cf. section 6.3.1.1), the mobility scenario is addressed in relation to information on air quality and noise and the use of internet of things.

3.3. Tourism

As tourism is a highly competitive domain, local tourism authorities have developed Web sites in order to offer of services to tourists. Unfortunately, the way information is organised does not necessarily meet Internet users expectations and numerous improvements are necessary to enhance their understanding of visited sites. Thus, even if only for economical reasons, the quality and the diversity of tourism packages have to be improved, for example by highlighting cultural heritage.

Again to illustrate our role in such a domain, let us cite some past projects where AxIS is involved related mainly to **Semantic Web Mining**[3]. In our case, a) we exploit ontologies and semantic data for improving usage analysis, personnalised services, the quality of results of search engines and for checking the content of an IS and also b) we exploit usage data for updating ontologies.) and Information Retrieval.

- Research has been carried out using log files from the city of Metz. This city was chosen because its Web site is in constant development and has been awarded several times, notably in 2003, 2004 and 2005 in the context of the Internet City label. The objective was to extract information about tourists behaviours from this site log files and to identify possible benefits in designing or updating a tourism ontology.
- Providing Tourism Information linked to Transportation information: AxIS has already studied recommender systems in order to provide users with personalised transportation information while looking for tourism information such as cultural information, leisure etc (cf. our recommender Be-TRIP (2006) based on CBR*tools).
- In the context of HOTEL-REF-PACA project (cf. section 6.1.2 , we aimed to better refer the web sites of hotels/campings from the region of TOURVAL in PACA (mainly Vésubie territory), with an approach based on a better understanding of usage from the internauts. To address this, we proposed and adopted a multidisicplinary approach combining various AxIs know-how: knowledge engineering (ontology in tourism), data mining (analysis of Google logs, hotel web site logs and user queries, visual behaviours from eye tracker), Ergonomics (clustering of hotel web sites based on their ergonomical quality).
- Several contacts (PACA, France Living Labs, Island of the Reunion) have been done related to projects in tourism and eco-tourism.

[3] By Semantic Web Mining, we mean the mutual benefits between two communities Semantic Web and Web Mining

3.4. User Involvement in Energy, Environment, Health and E-governement

Below are some topics where AxIS was or is involved in:

- **Preprocessing and analysing collective usage data and social networks** from group discussions related to design process: cf. ANR Intermed (2009) and FP7 Elliot (cf. section 6.3.1.1) where citizen generate ideas in terms of specific environmental sensors based services according to their needs.
- **Methods and tools for supporting open innovation based on public data**: a first work was made in 2010 with the CDISOD Color action related Public Data in collaboration with Fing (Marseille) and ADEME (Sophia Antipolis). We pursue such a study in the context of FP7 Elliot by providing to citizen environnemental data (air quality and noise) issued from citizen and/or territories sensors.

All AxIS topics are relevant for these domains. let us cite: social network analysis, personalization and information retrieval, recommender systems, expert search, design and evaluation of methods and tools for open innovation and user co-creation in the context of Living Labs, usage mining, mining data streams.

We have addressed specific works:

- Energy (cf. section 6.1.3): the main AxIS topic here was usage analysis in the context of an energy challenge in an enterprise (ECOFFICES) taking into account the complex and real situation (installation fo more than 400 sensors, differences between the three concerned teams, differences between the offices). Such an analysis aims to correlate team/office energy consuming, team/office eco-responsible behaviours and team/office profile.
- Health (cf. section 3.4): Axis contributed in 2011 to a Living Lab characterisation in Health domain through the visit of several Living Labs, which operate in the domain of Health and Autonomy, and conducted interviews. This work was done in relation with the CGIET [4].
- E-gov: The future Internet will bring a growing number of networked applications (services), devices and individual data (including private ones) to end-users. The important challenges are the organization of their access, and the guarantee of trust and privacy. The objectives of the PIMI [5] project (cf. section 6.2.1) are the definition of a design environment and a deployment platform for Personal Information Management system (PIM). The future PIM must provide the end-user personal data access with services that are relevant to his needs. In order to take mobility into account, the PIM will be accessed both by mobile devices (smartphone) and Personal Computers. With the increasing number of services and associated data being accessible through Internet, the number and complexity of PIM will augment dramatically in the near future. This will require strong research investment in a number of topics, all contributing to the expected usability and accessibility of Individual Information Spaces for the end-user.

[4] CGIET: "Conseil Général de l'Economie, de l'Industrie, de l'Energie et des technologies" http://www.cgeiet.economie.gouv.fr
[5] Personal Information Management through Internet

BANG Project-Team

4. Application Domains

4.1. Biology and medicine

The team is mostly involved in applications to biology and medicine. More precisely it aims at understanding biophysical mechanisms that sustain cell proliferation of malfunction. The main examples are biopolymers size repartition, cell self-organisation, tissu growth and cancer development or treatment.

4.2. Geophysical flows and environment

The team will split and give rise to another team ANGE specialised in complex geophysical flows in interaction with environment. Free surface flows as tsunamis, flows in river and costal areas and their ecological consequences are typical examples of applications developed in the team based on algorithms for the free-surface Navier-Stokes equations.

CAD Team

4. Application Domains

4.1. Introduction

Our scientific results in geometry have been tested in Aircraft industry, Ceramic industry, N-C simulation, and Computer Graphics as well.

4.2. Geometry

The cooperation with EADS, based on our new B-Spline surface formulation, is very promising, both for complex shape modelling and numerical simulations. Tolerance problems are currently studied in the sino-french Tsinghua PLM Center (supported by Dassault System).

Furthermore, this project also allowed to promoting several Associated Professors and Post Doctors in Chinese and French Research Institutes.

4.3. Computer Graphics

In Computer Graphics, our work in Computer Animation (fluid dynamics) has been tested in the Mr. Zhiyi Zhang's company.

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.

CLASSIC Project-Team

4. Application Domains

4.1. Forecasting of the electricity consumption

Our partner is EDF R&D. The goal is to aggregate in a sequential fashion the forecasts made by some (about 20) base experts in order to predict the electricity consumption at a global level (the one of all French customers) at a half-hourly step. We need to abide by some operational constraints: the predictions need to be made at noon for the next 24 hours (i.e., for the next 48 time rounds).

4.2. Forecasting of the air quality

Our partner is the Inria project-team CLIME (Paris-Rocquencourt). The goal is to aggregate in a sequential fashion the forecasts made by some (about 100) base experts in order to output field prediction of the concentration of some pollutants (typically, the ozone) over Europe. The results were and will be transferred to the public operator INERIS, which uses and will use them in an operational way.

4.3. Forecasting of the production data of oil reservoirs

Our partner is IFP Energies nouvelles. The goal is to aggregate in a sequential fashion the forecasts made by some (about 100) base experts in order to predict some behaviors (gas/oil ratio, cumulative oil extracted, water cut) of the exploitation of some oil wells.

4.4. Data mining, massive data sets

Our partner is the start-up Safety Line. The purpose of this application is to investigate statistical learning strategies for mining massive data sets originated from aircraft high-frequency recordings and improve security.

4.5. Computational linguistics

We propose and study new language models that bridge the gap between models oriented towards the statistical analysis of large corpora and grammars oriented towards the description of syntactic features as understood by academic experts. We have conceived a new kind of grammar, based on some cut and paste mechanism and some label aggregation principle, that can be fully learnt from a corpus. We are currently testing this model and studying its mathematical properties.

4.6. Statistical inference on biological data

The question is about understanding how interactions between neurons can be detected. A mathematical modeling is given by multivariate Hawkes processes. Lasso-type methods can then be used to estimate interaction functions in the nonparametric setting by using fast algorithms, providing inference of the unitary event activity of individual neurons.

CLIME Project-Team

4. Application Domains

4.1. Introduction

The central application domain of the project-team is atmospheric chemistry. We develop and maintain the air quality modeling system Polyphemus, which includes several numerical models (Gaussian models, Lagrangian model, two 3D Eulerian models including Polair3D) and their adjoints, and different high level methods: ensemble forecast, sequential and variational data assimilation algorithms. Advanced data assimilation methods, network design, inverse modeling, ensemble forecast are studied in the context of air chemistry. Note that addressing these high level issues requires controlling the full software chain (models and data assimilation algorithms).

The activity on assimilation of satellite data is mainly carried out for meteorology and oceanography. This is addressed in cooperation with external partners who provide numerical models. Concerning oceanography, the aim is to improve the forecast of ocean circulation, by assimilation of fronts and vortices displayed on image data. Concerning meteorology, the focus is on correcting the model location of structures related to high-impact weather events (cyclones, convective storms, *etc*.) by assimilating images.

4.2. Air quality

Air quality modeling implies studying the interactions between meteorology and atmospheric chemistry in the various phases of matter, which leads to the development of highly complex models. The different usages of these models comprise operational forecasting, case studies, impact studies, *etc*, with both societal (e.g., public information on pollution forecast) and economical impacts (e.g., impact studies for dangerous industrial sites). Models lack some appropriate data, for instance better emissions, to perform an accurate forecast and data assimilation techniques are recognized as a major key point for the improvement of forecast's quality.

In this context, Clime is interested in various problems, the following being the crucial ones:

- The development of ensemble forecast methods for estimating the quality of the prediction, in relation with the quality of the model and the observations. Sensitivity analysis with respect to the model's parameters so as to identify physical and chemical processes, whose modeling must be improved.
- The development of methodologies for sequential aggregation of ensemble simulations. What ensembles should be generated for that purpose, how spatialized forecasts can be generated with aggregation, how can the different approaches be coupled with data assimilation?
- The definition of second-order data assimilation methods for the design of optimal observation networks. Management of combinations of sensor types and deployment modes. Dynamic management of mobile sensors' trajectories.
- How to estimate the emission rate of an accidental release of a pollutant, using observations and a dispersion model (from the near-field to the continental scale)? How to optimally predict the evolution of a plume? Hence, how to help people in charge of risk evaluation for the population?
- The definition of non-Gaussian approaches for data assimilation.
- The assimilation of satellite measurements of troposphere chemistry.

The activities of Clime in air quality are supported by the development of the Polyphemus air quality modeling system. This system has a modular design, which makes it easier to manage high level applications such as inverse modeling, data assimilation and ensemble forecast.

4.3. Oceanography

The capacity of performing a high quality forecast of the state of the ocean, from the regional to the global scales, is of major interest. Such a forecast can only be obtained by systematically coupling numerical models and observations (*in situ* and satellite data). In this context, being able to assimilate image structures becomes a key point. Examples of such image structures are:

- apparent motion linked to surface velocity;
- trajectories, obtained either from tracking of features or from integration of the velocity field;
- spatial objects, such as fronts, eddies or filaments.

Image Models for these structures are developed and take into account the underlying physical processes. Image data are assimilated in Image Models to derive pseudo-observations of state variables, which are further assimilated in numerical ocean forecast models.

4.4. Meteorology

Meteorological forecasting constitutes a major applicative challenge for Image Assimilation. Although satellite data are operationally assimilated within models, this is mainly done on an independent pixel basis: the observed radiance is linked to the state variables via a radiative transfer model, that plays the role of an observation operator. Indeed, because of their limited spatial and temporal resolutions, numerical weather forecast models fail to exploit image structures, such as precursors of high impact weather:

- cyclogenesis related to the intrusion of dry stratospheric air in the troposphere (a precursor of cyclones),
- convective systems (supercells) leading to heavy winter time storms,
- low-level temperature inversion leading to fog and ice formation, *etc*.

To date, there is no available method for assimilating such data, which are characterized by a strong coherence in space and time. Meteorologists have developed qualitative Conceptual Models (CMs), for describing the high impact weathers and their signature on images, and tools to detect CMs on image data. The result of this detection is used for correcting the numerical models, for instance by modifying the initialization. The aim is therefore to develop a methodological framework allowing to assimilate the detected CMs within numerical forecast models. This is a challenging issue given the considerable impact of the related meteorological events.

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.

GAMMA3 Project-Team (section vide)

GANG Project-Team

3. Application Domains

3.1. Application Domains

Application domains include evaluating Internet performances, the design of new peer-to-peer applications, enabling large scale ad hoc networks and mapping the web.

- The application of measuring and modeling Internet metrics such as latencies and bandwidth is to provide tools for optimizing Internet applications. This concerns especially large scale applications such as web site mirroring and peer-to-peer applications.
- Peer-to-peer protocols are based on a all equal paradigm that allows to design highly reliable and scalable applications. Besides the file sharing application, peer-to-peer solutions could take over in web content dissemination resistant to high demand bursts or in mobility management. Envisioned peer-to-peer applications include video on demand, streaming, exchange of classified ads,...
- Wifi networks have entered our every day life. However, enabling them at large scale is still a challenge. Algorithmic breakthrough in large ad hoc networks would allow to use them in fast and economic deployment of new radio communication systems.
- The main application of the web graph structure consists in ranking pages. Enabling site level indexing and ranking is a possible application o f such studies.

HIPERCOM Project-Team

4. Application Domains

4.1. Introduction

The HIPERCOM project-team is mainly concerned by six domains:

- wireless mobile ad hoc networks,
- services over mobile networks,
- community networks,
- vehicular networks,
- large ad hoc networks with sensor nodes,
- energy-efficient wireless sensor networks.

4.2. Wireless mobile ad hoc networks

Wireless mobile ad hoc networks, Services over mobile networks, Community Networks.
Abstract. Mobile wireless networks have numerous applications in rescue and emergency operation, military tactical networking and in wireless high speed access to the internet.

A mobile ad hoc network is a network made of a collection of mobile nodes that gather spontaneously and communicate without requiring a pre-existing infrastructure. Of course a mobile ad hoc network use a wireless communication medium. They can be applied in various contexts:

- military;
- rescue and emergency;
- high speed access to internet.

The military context is the most obvious application of mobile ad hoc networks.
Soldiers invading a country won't susbscribe in advance to the local operator. On the reverse side, home units won't use their local operators firstly because they will likely be disrupted in the first hours of the conflict, and secondly because a wireless communication via an operator is not stealth enough to protect the data and the units. In Checheny, a general has been killed by a missile tracking the uplink signal of his portable phone.
The rescue context is halfway between military and civilian applications. In the september 11 disaster, most of the phone base station of the area have knocked out in less than twenty minutes. The remaining base stations were unable to operate because they could not work in ad hoc mode. The Wireless Emergency Rescue Team recommanded afterward that telecom operators should provide ad hoc mode for their infrastructure in order to operate in emergency situation in plain cooperation with police, firemen and hospital networks.
Mobile ad hoc network provide an enhanced coverage for high speed wireless access to the internet. The now very popular WLAN standard, WiFi, provides much larger capacity than mobile operator networks. Using a mobile ad hoc network around hot spots will offer high speed access to much larger community, including cars, busses, trains and pedestrians.

4.3. Services over mobile networks

Abstract. New wireless network calls for new services that fullfil the requirement in terms of mobility and capacity.

The generalization of a new generation of mobile networks calls for a new set of services and applications. For example:

- Indoor and outdoor positioning
- Service discovery and localisation
- Multicast and quality of services

Quality of service has become the central requirement that users expect from a network. High throughput, service continuity are critical issue for multimedia application over the wireless internet where the bandwidth is more scarce than in the wired world. A significant issue in the ad-hoc domain is that of the integrity of the network itself. Routing protocols allow, according to their specifications, any node to participate in the network - the assumption being that all nodes are behaving well and welcome. If that assumption fails - then the network may be subject to malicious nodes, and the integrity of the network fails. An important security service over mobile networks is to ensure that the integrity of the network is preserved even when attacks are launched against the integrity of the network.

4.4. Community Networks

Abstract. There is an increasing demand to deploy network within a community, rural or urban, with cabled or wireless access.

Community networks or citizen network are now frequent in big cities. In America most of the main cities have a community network. A community network is using the communication resource of each member (ADSL, Cable and wireless) to provide a general coverage of a city. Pedestrian in the street or in city mails can communicate via a high speed mobile mesh network. This new trend now appears in Europe with many experiments of the OLSR routing protocol in Paris, Lille, Toulouse, Berlin, Bruxelles, Seattle. The management of such networks is completely distributed and makes them very robust to faults. There is room for smart operators in this business.

4.5. Vehicular Networks

Abstract. Intelligent transport systems require efficient wireless telecommunications.

Vehicular ad hoc networks (VANET) are based on short- to medium-range transmission systems that support both vehicle-to-vehicle and vehicle-to-roadside communications. Vehicular networks will enable vehicular safety applications (safety warnings) as well as non-safety applications (real-time traffic information, routing support, mobile entertainment, and many others). We are interested in developing an efficient routing protocol that takes advantage of the fixed network infrastructure deployed along the roads. We are also studying MAC layer issues in order to provide more priority for security messages which have stringent delivery constraints.

4.6. Large ad hoc networks with sensor nodes

Abstract. Large autonomous wireless sensors in the internet of the things need very well tuned algorithms.

Self-organization is considered as a key element in tomorrow's Internet architecture. A major challenge concerning the integration of self-organized networks in the Internet is the accomplishment of light weight network protocols in large ad hoc environments.

In this domain, Hipercom's activity with wireless sensor nodes in collaboration with the Freie Universitaet in Berlin explores various solutions, including extensions of OLSR (for example DHT-OLSR) using programmable sensor nodes co-designed by the Freie Universitaet, and provides one of the largest testbeds of this kind, to date.

4.7. Energy efficient wireless sensor networks

Abstract. Energy efficiency is a key property in wireless sensor networks.

Various techniques are used to contribute to energy efficiency. In the OCARI network, an industrial wireless sensor network, we have designed and implemented an energy efficient routing protocol and a node activity scheduling algorithm allowing router nodes to sleep. We have applied a cross-layering approach allowing the optimization of MAC and network protocols taking into account the application requirements and the environment in which the network operates. This activity has been done in collaboration with our partners EDF, LIMOS and TELIT.

IMARA Project-Team

4. Application Domains

4.1. Introduction

While the preceding section focused on methodology, in connection with automated guided vehicles, it should be stressed that the evolution of the problems which we deal with remains often guided by the technological developments. We enumerate three fields of application, whose relative importance varies with time and who have strong mutual dependencies: driving assistance, cars available in self-service mode and fully automated vehicles (cybercars).

4.2. Driving assistance

Several techniques will soon help drivers. One of the first immediate goal is to improve security by alerting the driver when some potentially dangerous or dangerous situations arise, i.e. collision warning systems or lane tracking could help a bus driver and surrounding vehicle drivers to more efficiently operate their vehicles. Human factors issues could be addressed to control the driver workload based on additional information processing requirements.

Another issue is to optimize individual journeys. This means developing software for calculating optimal (for the user or for the community) path. Nowadays, path planning software is based on a static view of the traffic: efforts have to be done to take the dynamic component in account.

4.3. New transportation systems

The problems related to the abusive use of the individual car in large cities led the populations and the political leaders to support the development of public transport. A demand exists for a transport of people and goods which associates quality of service, environmental protection and access to the greatest number. Thus the tram and the light subways of VAL type recently introduced into several cities in France conquered the populations, in spite of high financial costs.

However, these means of mass transportation are only possible on lines on which there is a keen demand. As soon as one moves away from these “lines of desire” or when one deviates from the rush hours, these modes become expensive and offer can thus only be limited in space and time.

To give a more flexible offer, it is necessary to plan more individual modes which approach the car as we know it. However, if one wants to enjoy the benefits of the individual car without suffering from their disadvantages, it is necessary to try to match several criteria: availability anywhere and anytime to all, lower air and soils pollution as well as sound levels, reduced ground space occupation, security, low cost.

Electric or gas vehicles available in self-service as in the Praxitèle system bring a first response to these criteria. To be able to still better meet the needs, it is however necessary to re-examine the design of the vehicles on the following points:

- ease empty car moves to better distribute them;
- better use of information systems inboard and on ground;
- better integrate this system in the global transportation system.

These systems are now operating (i.e. in La Rochelle). The challenge is to bring them to an industrial phase by transferring technologies to these still experimental projects.

4.4. Cybercars

The long term effort of the project is to put automatically guided vehicles (cybercars) on the road. It seems too early to mix cybercars and traditional vehicles, but data processing and automation now make it possible to consider in the relatively short term the development of such vehicles and the adapted infrastructures. IMARA aims at using these technologies on experimental platforms (vehicles and infrastructures) to accelerate the technology transfer and to innovate in this field.

Other application can be precision docking systems that will allow buses to be automatically maneuvered into a loading zone or maintenance area, allowing easier access for passengers, or more efficient maintenance operations. Transit operating costs will also be reduced through decreased maintenance costs and less damage to the breaking and steering systems.

Regarding technical topics, several aspects of Cybercars have been developed at IMARA this year. First, we have stabilized a generic Cycab architecture involving Inria Syndex tool and CAN communications. The critical part of the vehicle is using a real time Syndex application controlling the actuators via two Motorola's MPC555. Today, we have decided to migrate to the new dsPIC architecture for more efficiency and ease of use.

This application has a second feature, it can receive commands from an external source (Asynchronously this time) on a second CAN bus. This external source can be a PC or a dedicated CPU, we call it high level. To work on the high level, in the past years we have been developing a R&D framework called (Taxi) which used to take control of the vehicle (Cycab and Yamaha) and process data such as gyro, GPS, cameras, wireless communications and so on. Today, in order to rely on a professional and maintained solution, we have chosen to migrate to the RTMAPS SDK development platform. Today, all our developments and demonstrations are using this efficient prototyping platform. Thanks to RT-MAPS we've been able to do all the demonstrations on our cybercars: cycabs, Yamaha AGV and new Cybus platforms. These demonstrations include: reliable SLAMMOT algorithm using 2 to 4 laser sensors simultaneously, automatic line/road following techniques, PDA remote control, multi sensors data fusion, collaborative perception via ad-hoc network.

The second main topic is inter-vehicle communications using ad-hoc networks. We have worked with the HIPERCOM team for setting and tuning OLSR, a dynamic routing protocol for vehicles communications (see Section 3.2). Our goal is to develop a vehicle dedicated communication software suite, running on a specialized hardware. It can be linked also with the Taxi Framework for getting data such GPS information's to help the routing algorithm.

IMEDIA2 Team

4. Application Domains

4.1. Scientific applications

Examples: environmental images databases: fauna and flora; satellite images databases: ground typology; medical images databases: find images of a pathological character for educational or investigation purposes.
We are developing tools enabling multimedia access to biodiversity collections for species identifications inside the Pl@ntNet project.
In fact, almost all IMEDIA2 team members are involved in the Pl@ntNet project and develop methods that can be used in this context.

4.2. Audio-visual applications

Examples: Look for a specific shot in a movie, documentary or TV news, present a video summary. Help archivists to annotate the contents. Retrieve copies of a given material (photo or video) in a TV stream or on the web.
Our team has a collaboration with AFP press agency in the context of GLOCAL project.

MATHRISK Team

3. Application Domains

3.1. Application Domains

- Utility maximization in incomplete markets
- option pricing
- quantitative risk management
- systemic risk
- limit order books
- credit risk
- liquidity risk
- computational finance
- calibration

MICMAC Project-Team

4. Application Domains

4.1. Electronic structure of large systems

As the size of the systems one wants to study increases, more efficient numerical techniques need to be resorted to. In computational chemistry, the typical scaling law for the complexity of computations with respect to the size of the system under study is N^3, N being for instance the number of electrons. The Holy Grail in this respect is to reach a linear scaling, so as to make possible simulations of systems of practical interest in biology or material science. Efforts in this direction must address a large variety of questions such as

- how can one improve the nonlinear iterations that are the basis of any *ab initio* models for computational chemistry?
- how can one more efficiently solve the inner loop which most often consists in the solution procedure for the linear problem (with frozen nonlinearity)?
- how can one design a sufficiently small variational space, whose dimension is kept limited while the size of the system increases?

An alternative strategy to reduce the complexity of *ab initio* computations is to try to couple different models at different scales. Such a mixed strategy can be either a sequential one or a parallel one, in the sense that

- in the former, the results of the model at the lower scale are simply used to evaluate some parameters that are inserted in the model for the larger scale: one example is the parameterized classical molecular dynamics, which makes use of force fields that are fitted to calculations at the quantum level;
- while in the latter, the model at the lower scale is concurrently coupled to the model at the larger scale: an instance of such a strategy is the so called QM/MM coupling (standing for Quantum Mechanics/Molecular Mechanics coupling) where some part of the system (typically the reactive site of a protein) is modeled with quantum models, that therefore accounts for the change in the electronic structure and for the modification of chemical bonds, while the rest of the system (typically the inert part of a protein) is coarse grained and more crudely modeled by classical mechanics.

The coupling of different scales can even go up to the macroscopic scale, with methods that couple a microscopic description of matter, or at least a mesoscopic one, with the equations of continuum mechanics at the macroscopic level.

4.2. Computational Statistical Mechanics

The orders of magnitude used in the microscopic description of matter are far from the orders of magnitude of the macroscopic quantities we are used to: The number of particles under consideration in a macroscopic sample of material is of the order of the Avogadro number $\mathcal{N}_A \sim 10^{23}$, the typical distances are expressed in Å (10^{-10} m), the energies are of the order of $k_{\mathrm{B}}T \simeq 4 \times 10^{-21}$ J at room temperature, and the typical times are of the order of 10^{-15} s when the proton mass is the reference mass.

To give some insight into such a large number of particles contained in a macroscopic sample, it is helpful to compute the number of moles of water on earth. Recall that one mole of water corresponds to 18 mL, so that a standard glass of water contains roughly 10 moles, and a typical bathtub contains 10^5 mol. On the other hand, there are approximately 1.3×10^{18} m^3 of water in the oceans, *i.e.* 7.2×10^{22} mol, a number comparable to the Avogadro number. This means that inferring the macroscopic behavior of physical systems described at the microscopic level by the dynamics of several millions of particles only is like inferring the ocean's dynamics from hydrodynamics in a bathtub...

For practical numerical computations of matter at the microscopic level, following the dynamics of every atom would require simulating $\mathcal{N}_A$ atoms and performing $\mathrm{O}(10^{15})$ time integration steps, which is of course impossible! These numbers should be compared with the current orders of magnitude of the problems that can be tackled with classical molecular simulation, where several millions of atoms only can be followed over time scales of the order of 0.1 μs.

Describing the macroscopic behavior of matter knowing its microscopic description therefore seems out of reach. Statistical physics allows us to bridge the gap between microscopic and macroscopic descriptions of matter, at least on a conceptual level. The question is whether the estimated quantities for a system of N particles correctly approximate the macroscopic property, formally obtained in the thermodynamic limit $N \to +\infty$ (the density being kept fixed). In some cases, in particular for simple homogeneous systems, the macroscopic behavior is well approximated from small-scale simulations. However, the convergence of the estimated quantities as a function of the number of particles involved in the simulation should be checked in all cases.

Despite its intrinsic limitations on spatial and timescales, molecular simulation has been used and developed over the past 50 years, and its number of users keeps increasing. As we understand it, it has two major aims nowadays.

First, it can be used as a *numerical microscope*, which allows us to perform "computer" experiments. This was the initial motivation for simulations at the microscopic level: physical theories were tested on computers. This use of molecular simulation is particularly clear in its historic development, which was triggered and sustained by the physics of simple liquids. Indeed, there was no good analytical theory for these systems, and the observation of computer trajectories was very helpful to guide the physicists' intuition about what was happening in the system, for instance the mechanisms leading to molecular diffusion. In particular, the pioneering works on Monte-Carlo methods by Metropolis et al, and the first molecular dynamics simulation of Alder and Wainwright were performed because of such motivations. Today, understanding the behavior of matter at the microscopic level can still be difficult from an experimental viewpoint (because of the high resolution required, both in time and in space), or because we simply do not know what to look for! Numerical simulations are then a valuable tool to test some ideas or obtain some data to process and analyze in order to help assessing experimental setups. This is particularly true for current nanoscale systems.

Another major aim of molecular simulation, maybe even more important than the previous one, is to compute macroscopic quantities or thermodynamic properties, typically through averages of some functionals of the system. In this case, molecular simulation is a way to obtain *quantitative* information on a system, instead of resorting to approximate theories, constructed for simplified models, and giving only qualitative answers. Sometimes, these properties are accessible through experiments, but in some cases only numerical computations are possible since experiments may be unfeasible or too costly (for instance, when high pressure or large temperature regimes are considered, or when studying materials not yet synthesized). More generally, molecular simulation is a tool to explore the links between the microscopic and macroscopic properties of a material, allowing one to address modelling questions such as "Which microscopic ingredients are necessary (and which are not) to observe a given macroscopic behavior?"

4.3. Homogenization and related problems

Over the years, the project-team has developed an increasing expertise on how to couple models written at the atomistic scale, with more macroscopic models, and, more generally, an expertise in multiscale modelling for materials science.

The following observation motivates the idea of coupling atomistic and continuum description of materials. In many situations of interest (crack propagation, presence of defects in the atomistic lattice, ...), using a model based on continuum mechanics is difficult. Indeed, such a model is based on a macroscopic constitutive law, the derivation of which requires a deep qualitative and quantitative understanding of the physical and mechanical properties of the solid under consideration. For many solids, reaching such an understanding is a challenge, as loads they are submitted to become larger and more diverse, and as experimental observations

helping designing such models are not always possible (think of materials used in the nuclear industry). Using an atomistic model in the whole domain is not possible either, due to its prohibitive computational cost. Recall indeed that a macroscopic sample of matter contains a number of atoms on the order of 10^{23}. However, it turns out that, in many situations of interest, the deformation that we are after is not smooth in *only a small part* of the solid. So, a natural idea is to try to take advantage of both models, the continuum mechanics one and the atomistic one, and to couple them, in a domain decomposition spirit. In most of the domain, the deformation is expected to be smooth, and reliable continuum mechanics models are then available. In the rest of the domain, the expected deformation is singular, one needs an atomistic model to describe it properly, the cost of which remains however limited as this region is small.

From a mathematical viewpoint, the question is to couple a discrete model with a model described by PDEs. This raises many questions, both from the theoretical and numerical viewpoints:

- first, one needs to derive, from an atomistic model, continuum mechanics models, under some regularity assumptions that encode the fact that the situation is smooth enough for such a macroscopic model to be a good description of the materials;
- second, couple these two models, e.g. in a domain decomposition spirit, with the specificity that models in both domains are written in a different language, that there is no natural way to write boundary conditions coupling these two models, and that one would like the decomposition to be self-adaptive.

More generally, the presence of numerous length-scales in material science problems represents a challenge for numerical simulation, especially when some *randomness* is assumed on the materials. It can take various forms, and includes defects in crystals, thermal fluctuations, and impurities or heterogeneities in continuous media. Standard methods available in the literature to handle such problems often lead to very costly computations. Our goal is to develop numerical methods that are more affordable. Because we cannot embrace all difficulties at once, we focus on a simple case, where the fine scale and the coarse-scale models can be written similarly, in the form of a simple elliptic partial differential equation in divergence form. The fine scale model includes heterogeneities at a small scale, a situation which is formalized by the fact that the coefficients in the fine scale model vary on a small length scale. After homogenization, this model yields an effective, macroscopic model, which includes no small scale. In many cases, a sound theoretical groundwork exists for such homogenization results. We consider mostly the setting of stochastic homogenization of linear, scalar, second order elliptic PDEs, where analytical formulas for the effective properties are known. The difficulty stems from the fact that they generally lead to prohibitively costly computations. For such a case, simple from the theoretical viewpoint, our aim is to focus on different practical computational approaches to speed-up the computations. One possibility, among others, is to look for specific random materials, relevant from the practical viewpoint, and for which a dedicated approach can be proposed, that is less expensive than the general approach.

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.

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].

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.

POMDAPI Project-Team (section vide)

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.

RAP Project-Team (section vide)

REGAL Project-Team

4. Application Domains

4.1. Research domain

To address the evolution of distributed platforms in recent years, we focus on the following areas:

- *Distributed algorithms for dynamic and large networks.* Network topology is no more static; distributed systems are increasingly dynamic, i.e., nodes can join, fail, recover, disconnect and reconnect, and change location. Examples include IaaS cloud computing infrastructures, where virtual machines can be moved according to load peaks, opportunistic networks such as DTNs (Delay-Tolerant Networks), and networks of robots.
- *Management of distributed data.* In emerging architectures such as distributed hash tables (DHTs) and cloud computing, our research topics include replica placement, responsiveness, load balancing, consistency maintenance, consensus algorithms, and synchronisation. This research direction is funded by several new collaborative projects (ConcoRDanT, MyCloud, Nu@age, Odisea, Prose, Shaman, Spades, Streams, R-Discover) and by industrial funding (Google).
- *Performance and robustness of Systems Software in multicore architectures.* Our research focuses on the efficient management of system resources at the user level. Issues considered include efficient synchronization and memory management in large-scale multicore architectures. At the same time, we focus on the robustness of systems software, based on the Coccinelle technology. This work is funded by ANR ABL and InfraJVM.

REO Project-Team

4. Application Domains

4.1. Blood flows

Cardiovascular diseases like atherosclerosis or aneurysms are a major cause of mortality. It is generally admitted that a better knowledge of local flow patterns could improve the treatment of these pathologies (although many other biophysical phenomena obviously take place in the development of such diseases). In particular, it has been known for years that the association of low wall shear stress and high oscillatory shear index give relevant indications to localize possible zones of atherosclerosis. It is also known that medical devices (graft or stent) perturb blood flows and may create local stresses favorable with atherogenesis. Numerical simulations of blood flows can give access to this local quantities and may therefore help to design new medical devices with less negative impacts. In the case of aneurysms, numerical simulations may help to predict possible zones of rupture and could therefore give a guide for treatment planning.

In clinical routine, many indices are used for diagnosis. For example, the size of a stenosis is estimated by a few measures of flow rate around the stenosis and by application of simple fluid mechanics rules. In some situations, for example in the case a sub-valvular stenosis, it is known that such indices often give false estimations. Numerical simulations may give indications to define new indices, simple enough to be used in clinical exams, but more precise than those currently used.

It is well-known that the arterial circulation and the heart (or more specifically the left ventricle) are strongly coupled. Modifications of arterial walls or blood flows may indeed affect the mechanical properties of the left ventricle. Numerical simulations of the arterial tree coupled to the heart model could shed light on this complex relationship.

One of the goals of the REO team is to provide various models and simulation tools of the cardiovascular system. The scaling of these models will be adapted to the application in mind: low resolution for modeling the global circulation, high resolution for modeling a small portion of vessel.

4.2. Respiratory tracts

Breathing, or “external” respiration (“internal” respiration corresponds to cellular respiration) involves gas transport though the respiratory tract with its visible ends, nose and mouth. Air streams then from the pharynx down to the trachea. Food and drink entry into the trachea is usually prevented by the larynx structure (epiglottis). The trachea extends from the neck into the thorax, where it divides into right and left main bronchi, which enter the corresponding lungs (the left being smaller to accommodate the heart). Inhaled air is then convected in the bronchus tree which ends in alveoli, where gaseous exchange occurs. Surfactant reduces the surface tension on the alveolus wall, allowing them to expand. Gaseous exchange relies on simple diffusion on a large surface area over a short path between the alveolus and the blood capillary under concentration gradients between alveolar air and blood. The lungs are divided into lobes (three on the right, two on the left) supplied by lobar bronchi. Each lobe of the lung is further divided into segments (ten segments of the right lung and eight of the left). Inhaled air contains dust and debris, which must be filtered, if possible, before they reach the alveoli. The tracheobronchial tree is lined by a layer of sticky mucus, secreted by the epithelium. Particles which hit the side wall of the tract are trapped in this mucus. Cilia on the epithelial cells move the mucous continually towards the nose and mouth.

Each lung is enclosed in a space bounded below by the diaphragm and laterally by the chest wall and the mediastinum. The air movement is achieved by alternately increasing and decreasing the chest pressure (and volume). When the airspace transmural pressure rises, air is sucked in. When it decreases, airspaces collapse and air is expelled. Each lung is surrounded by a pleural cavity, except at its hilum where the inner pleura give birth to the outer pleura. The pleural layers slide over each other. The tidal volume is nearly equal to 500 ml.

The lungs may fail to maintain an adequate supply of air. In premature infants surfactant is not yet active. Accidental inhalation of liquid or solid and airway infection may occur. Chronic obstructive lung diseases and lung cancers are frequent pathologies and among the three first death causes in France.

One of the goals of REO team in the ventilation field is to visualize the airways (virtual endoscopy) and simulate flow in image-based 3D models of the upper airways (nose, pharynx, larynx) and the first generations of the tracheobronchial tree (trachea is generation 0), whereas simple models of the small bronchi and alveoli are used (reduced-basis element method, fractal homogenization, multiphysics homogenization, lumped parameter models), in order to provide the flow distribution within the lung segments. This activity has been carried out in the framework of successive research programs: RNTS “R-MOD” until 2005, ACI “le-poumon-vous-dis-je” until 2007 and ANR M3RS until 2013.

4.3. Cardiac electrophysiology

The purpose is to simulate the propagation of the action potential in the heart. A lot of works has already been devoted to this topic in the literature (see *e.g.* [71], [75], [74] and the references therein), nevertheless there are only very few studies showing realistic electrocardiograms obtained from partial differential equations models. Our goal is to find a compromise between two opposite requirements: on the one hand, we want to use predictive models, and therefore models based on physiology, on the other hand, we want to use models simple enough to be parametrized (in view of patient-specific simulations). We are now working on using our ECG simulator to address the inverse problem of electrocardiology. In collaboration with the Macsproject-team, we are working on the electromechanical coupling in the myocardium. We are also interested in various clinical and industrial issues related to cardiac electrophysiology. In particular, we collaborated with ELA Medical company (pacemaker manufacturer, Sorin group).

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.

SIERRA Project-Team

4. Application Domains

4.1. Application Domains

Machine learning research can be conducted from two main perspectives: the first one, which has been dominant in the last 30 years, is to design learning algorithms and theories which are as generic as possible, the goal being to make as few assumptions as possible regarding the problems to be solved and to let data speak for themselves. This has led to many interesting methodological developments and successful applications. However, we believe that this strategy has reached its limit for many application domains, such as computer vision, bioinformatics, neuro-imaging, text and audio processing, which leads to the second perspective our team is built on: Research in machine learning theory and algorithms should be driven by interdisciplinary collaborations, so that specific prior knowledge may be properly introduced into the learning process, in particular with the following fields:

- Computer vision: objet recognition, object detection, image segmentation, image/video processing, computational photography. In collaboration with the Willow project-team.
- Bioinformatics: cancer diagnosis, protein function prediction, virtual screening. In collaboration with Institut Curie.
- Text processing: document collection modeling, language models.
- Audio processing: source separation, speech/music processing. In collaboration with Telecom Paristech.
- Neuro-imaging: brain-computer interface (fMRI, EEG, MEG). In collaboration with the Parietal project-team.

SISYPHE Project-Team (section vide)

SMIS Project-Team

4. Application Domains

4.1. Application Domains

data privacy, personal information management, healthcare, ambient intelligence

Our work addresses varied application domains. Typically, data management techniques on chip are required each time data-driven applications have to be embedded in ultra-light computing devices. This situation occurs for example in healthcare applications where medical folders are embedded into smart tokens (e.g., smart cards, secured USB keys), in telephony applications where personal data (address book, agenda, etc.) is embedded into cellular phones, in sensor networks where sensors log row measurements and perform local computation on them, in smart-home applications where a collection of smart appliances gather information about the occupants to provide them a personalized service, and more generally in most applications related to ambient intelligence.

Safeguarding data confidentiality has become a primary concern for citizens, administrations and companies, broadening the application domains of our work on access control policies definition and enforcement. The threat on data confidentiality is manifold: external and internal attacks on the data at rest, on the data on transit, on the data hosted in untrusted environments (e.g., Database Service Providers, Web-hosting companies) and subject to illegal usage, insidious gathering of personal data in an ambient intelligence surrounding. Hence, new access control models and security mechanisms are required to accurately declare and safely control who is granted access to which data and for which purpose.

While the application domain mentioned above is rather large, one application is today more specifically targeted by the SMIS project. This application deals with privacy preservation in EHR (Electronic Health Record) systems. Several countries (including France) launched recently ambitious EHR programs where medical folders will be centralized and potentially hosted by private Database Service Providers. Centralization and hosting increase the risk of privacy violation. In 2007, we launched two projects (PlugDB and DMSP) tackling precisely this issue, with the final objective to experiment our technologies in the field. In 2011, we launched a new project (KISS) capitalizing on the previous ones and extending their scope towards the protection of any personal data delivered to individuals in an electronic form.

TREC Project-Team

4. Application Domains

4.1. Application Domains

We have investigated various applications of our research results with the following industrial partners and user associations:

- **Wireless Networks**
  - Alcatel-Lucent Bell Laboratories (L. Thomas and L. Roullet) on self optimization in cellular networks.
  - Qualcomm (T. Richardson and his group) on improvements of CSMA CA.
  - Orange (M. Karray) on cellular networks.
- **Network Dynamics**
  - Thalès and Real-Time-at-Work on embedded networks.
  - Grenouille on probing in access networks.
- **Networks Economics**
  - Technicolor (J. Bolot) on economic incentives.

WILLOW Project-Team

4. Application Domains

4.1. Introduction

We believe that foundational modeling work should be grounded in applications. This includes (but is not restricted to) the following high-impact domains.

4.2. Quantitative image analysis in science and humanities

We plan to apply our 3D object and scene modeling and analysis technology to image-based modeling of human skeletons and artifacts in anthropology, and large-scale site indexing, modeling, and retrieval in archaeology and cultural heritage preservation. Most existing work in this domain concentrates on image-based rendering—that is, the synthesis of good-looking pictures of artifacts and digs. We plan to focus instead on quantitative applications. We are engaged in a project involving the archaeology laboratory at ENS and focusing on image-based artifact modeling and decorative pattern retrieval in Pompeii. This effort is part of the MSR-Inria project mentioned earlier and that will be discussed further later in this report. Application of our 3D reconstruction technology is now being explored in the field of cultural heritage and archeology by the start-up Iconem, founded by Y. Ubelmann, a Willow collaborator.

4.3. Video Annotation, Interpretation, and Retrieval

Both specific and category-level object and scene recognition can be used to annotate, augment, index, and retrieve video segments in the audiovisual domain. The Video Google system developed by Sivic and Zisserman (2005) for retrieving shots containing specific objects is an early success in that area. A sample application, suggested by discussions with Institut National de l'Audiovisuel (INA) staff, is to match set photographs with actual shots in film and video archives, despite the fact that detailed timetables and/or annotations are typically not available for either medium. Automatically annotating the shots is of course also relevant for archives that may record hundreds of thousands of hours of video. Some of these applications will be pursued in our MSR-Inria project, in which INA is one of our partners.