RESEARCH CENTER

FIELD
Algorithmics, Programming, Software and Architecture

Activity Report 2017

Section New Results

Edition: 2018-02-19

ARIC Project-Team

7. New Results

7.1. Efficient approximation methods

7.1.1. Automatic generation of hardware FIR filters from a frequency domain specification

In [53], we present an open-source tool for the automatic design of reliable finite impulse response (FIR) filters, targeting FPGAs. It shows that user intervention can be limited to a very small number of relevant input parameters: a high-level frequency-domain specification, and input/output formats. All the other design parameters are computed automatically, using novel approaches to filter coefficient quantization and direct-form architecture implementation. Our tool guarantees a priori that the resulting architecture respects the specification while attempting to minimize its cost. Our approach is evaluated on a range of examples and shown to produce designs that are very competitive with the state of the art, with very little design effort.

7.1.2. Exponential sums and correctly-rounded functions

The 2008 revision of the IEEE-754 standard, which governs floating-point arithmetic, recommends that a certain set of elementary functions should be correctly rounded. Successful attempts for solving the Table Maker's Dilemma in binary64 made it possible to design `CRlibm`, a library which offers correctly rounded evaluation in binary64 of some functions of the usual `libm`. It evaluates functions using a two step strategy, which relies on a folklore heuristic that is well spread in the community of mathematical functions designers. Under this heuristic, one can compute the distribution of the lengths of runs of zeros/ones after the rounding bit of the value of the function at a given floating-point number. The goal of [13] was to change, whenever possible, this heuristic into a rigorous statement. The underlying mathematical problem amounts to counting integer points in the neighborhood of a curve, which we tackle using so-called exponential sums techniques, a tool from analytic number theory.

7.1.3. Continued fractions in power series fields

In [5], we explicitly describe a noteworthy transcendental continued fraction in the field of power series over $\mathbb{Q}$, having irrationality measure equal to 3. This continued fraction is a generating function of a particular sequence in the set $\{1, 2\}$. The origin of this sequence, whose study was initiated in a recent paper, is to be found in another continued fraction, in the field of power series over $\mathbb{F}_3$, which satisfies a simple algebraic equation of degree 4, introduced thirty years ago by D. Robbins.

7.1.4. Validated and numerically efficient Chebyshev spectral methods for linear ordinary differential equations

In [51], we develop a validated numerics method for the solution of linear ordinary differential equations (LODEs). A wide range of algorithms (i.e., Runge-Kutta, collocation, spectral methods) exist for numerically computing approximations of the solutions. Most of these come with proofs of asymptotic convergence, but usually, provided error bounds are non-constructive. However, in some domains like critical systems and computer-aided mathematical proofs, one needs validated effective error bounds. We focus on both the theoretical and practical complexity analysis of a so-called *a posteriori* quasi-Newton validation method, which mainly relies on a fixed-point argument of a contracting map. Specifically, given a polynomial approximation, obtained by some numerical algorithm and expressed in Chebyshev basis, our algorithm efficiently computes an accurate and rigorous error bound. For this, we study theoretical properties like compactness, convergence, invertibility of associated linear integral operators and their truncations in a suitable coefficient space of Chebyshev series. Then, we analyze the almost-banded matrix structure of these operators, which allows for very efficient numerical algorithms for both numerical solutions of LODEs and rigorous computation of the approximation error. Finally, several representative examples show the advantages of our algorithms as well as their theoretical and practical limits.

7.1.5. Validated semi-analytical transition matrices for linearized relative spacecraft dynamics via Chebyshev series appproximations

In [47], we provide an efficient generic algorithm to compute validated approximations of transition matrices of linear time-variant systems using Chebyshev expansions, and apply it to two different examples of relative motion of satellites (spacecraft rendezvous with Tschauner-Hempel equations and geostationary station keeping with J2 perturbation in the linearized Orange model).

7.2. Lattices: algorithms and cryptology

7.2.1. All-But-Many Lossy Trapdoor Functions and Selective Opening Chosen-Ciphertext Security from LWE

In cryptography, selective opening (SO) security refers to adversaries that receive a number of ciphertexts and, after having corrupted a subset of the senders (thus obtaining the plaintexts and the senders' random coins), aim at breaking the security of remaining ciphertexts. So far, very few public-key encryption schemes are known to provide simulation-based selective opening (SIM-SO-CCA2) security under chosen-ciphertext attacks and most of them encrypt messages bit-wise. The only exceptions to date rely on all-but-many lossy trapdoor functions (as introduced by Hofheinz; Eurocrypt'12) and the Composite Residuosity assumption. In a paper [43] published at Crypto 2017, the team describes the first all-but-many lossy trapdoor function with security relying on the presumed hardness of the Learning-With-Errors problem (LWE) with standard parameters. The new construction exploits homomorphic computations on lattice trapdoors for lossy LWE matrices. By carefully embedding a lattice trapdoor in lossy public keys, the paper is able to prove SIM-SO-CCA2 security under the LWE assumption. As a result of independent interest, the paper describes a variant of our scheme whose multi-challenge CCA2 security tightly relates to the hardness of LWE and the security of a pseudo-random function.

7.2.2. Zero-Knowledge Arguments for Lattice-Based PRFs and Applications to E-Cash

This paper [41] deals with cryptographic pseudorandom functions from lattice assumptions and their use in e-cash systems. Beyond their security guarantees under well-studied assumptions, algebraic pseudo-random functions are motivated by their compatibility with efficient zero-knowledge proof systems, which is useful in a number of privacy applications like digital cash. The paper considers the problem of proving the correct evaluation of lattice-based PRFs based on the Learning-With-Rounding (LWR) problem introduced by Banerjee et al. (Eurocrypt'12). Namely, the paper provides zero-knowledge arguments of knowledge of triples (y, k, x) such that $y = F_k(x)$ is the correct evaluation of a PRF for a secret input x and a committed key k. While analogous statements admit efficient zero-knowledge protocols in the discrete logarithm setting, they have never been addressed in lattices so far. The paper provides such arguments for the key homomorphic PRF of Boneh et al. (Crypto'13) and the generic PRF implied by the LWR-based pseudo-random generator. As an application, the paper describes the first compact e-cash system based on lattice assumptions.

7.2.3. Adaptive Oblivious Transfer with Access Control from Lattice Assumptions

Adaptive oblivious transfer (OT) is a cryptographic protocol where a sender initially commits to a database $\{M_i\}_{i=1}^N$. Then, a receiver can query the sender up to k times with private indexes $\rho_1, ..., \rho_k$ so as to obtain $M_{\rho_1}, ..., M_{\rho_k}$ and nothing else. Moreover, for each $i \in [k]$, the receiver's choice ρ_i may depend on previously obtained messages. Oblivious transfer with access control (OT-AC) is a flavor of adaptive OT where database records are protected by distinct access control policies that specify which credentials a receiver should obtain in order to access each M_i. So far, all known OT-AC protocols only support access policies made of conjunctions or rely on *ad hoc* assumptions in pairing-friendly groups (or both). The paper [40] provides an OT-AC protocol where access policies may consist of any branching program of polynomial length, which is sufficient to realize any access policy in NC1. The security of the protocol is proved under the Learning-with-Errors (LWE) and Short-Integer-Solution (SIS) assumptions. As a result of independent interest, the paper provides protocols for proving the correct evaluation of a committed branching program on a committed input.

7.2.4. Encoding-Free ElGamal-Type Encryption Schemes on Elliptic Curves

At PKC 2006, Chevallier-Mames, Paillier, and Pointcheval proposed a very elegant technique over cyclic subgroups of $\mathbb{F}_p$ eliminating the need to encode the message as a group element in the ElGamal encryption scheme. Unfortunately, it is unclear how to adapt their scheme over elliptic curves. In a previous attempt, Virat suggested an adaptation of ElGamal to elliptic curves over the ring of dual numbers as a way to address the message encoding issue. Advantageously the resulting cryptosystem does not require encoding messages as points on an elliptic curve prior to their encryption. Unfortunately, it only provides one-wayness and, in particular, it is not (and was not claimed to be) semantically secure. The paper revisits Virat's cryptosystem and extends the Chevallier-Mames et al.'s technique to the elliptic curve setting. The paper [35] considers elliptic curves over the ring $\mathbb{Z}/(p^2\mathbb{Z})$ and defines the underlying class function. This yields complexity assumptions whereupon new ElGamal-type encryption schemes are built. The so-obtained schemes are proved semantically secure and make use of a very simple message encoding: messages being encrypted are viewed as elements in the range $[0, p-1]$. Further, the new schemes come equipped with a partial ring-homomorphism property: anyone can add a constant to an encrypted message –or– multiply an encrypted message by a constant. This can prove helpful as a blinding method in a number of applications. Finally, in addition to practicability, the proposed schemes also offer better performance in terms of speed, memory, and bandwidth.

7.2.5. Structure-Preserving Chosen-Ciphertext Security with Shorter Verifiable Ciphertexts

Structure-preserving cryptography is a world where messages, signatures, ciphertexts and public keys are entirely made of elements of a group over which a bilinear map is efficiently computable. This property makes the primitives compatible with the Groth-Sahai non-interactive proof systems in the design of higher-level privacy-preserving protocols. While structure-preserving signatures have received much attention the last 6 years, structure-preserving encryption schemes have undergone slower development. In particular, the best known structure-preserving cryptosystems with chosen-ciphertext (IND-CCA2) security either rely on symmetric pairings or require long ciphertexts comprised of hundreds of group elements or do not provide publicly verifiable ciphertexts. The paper [42] provides a publicly verifiable construction based on the SXDH assumption in asymmetric bilinear groups $e : G_1 \times G_2 \to G_T$, which features relatively short ciphertexts. For typical parameters, the ciphertext size amounts to less than 40 elements of G. As a second contribution, the paper provides a structure-preserving encryption scheme with perfectly randomizable ciphertexts and replayable chosen-ciphertext security. The new RCCA-secure system significantly improves upon the best known system featuring similar properties in terms of ciphertext size.

7.2.6. Tightly Secure IBE under Constant-size Master Public Key

This paper is about identity-based encryption (IBE). Chen and Wee (Crypto 2013) proposed the first almost tightly and adaptively secure IBE in the standard model and left two open problems which called for a tightly secure IBE with (1) constant-size master public key and/or (2) constant security loss. This paper proposes an IBE scheme with constant-size master public key and tighter security reduction. This (partially) solves Chen and Wee's first open problem and makes progress on the second one. Technically, the new IBE scheme is built based on Wee's petit IBE scheme (TCC 2016) in composite-order bilinear groups whose order is product of four primes. The sizes of master public key, ciphertexts, and secret keys are not only constant but also nearly optimal as Wee's petit IBE. The paper [33] proves its adaptive security in the multi-instance, multi-ciphertext setting (PKC 2015) based on the decisional subgroup assumption and a subgroup variant of DBDH assumption. The security loss is $O(\log q)$ where q is the upper bound of the total number of secret keys and challenge ciphertexts revealed to adversary in each single IBE instance. It is much smaller than those for all known adaptively secure IBE schemes in a concrete sense.

7.2.7. ABE with Tag Made Easy: Concise Framework and New Instantiations in Prime-order Groups

Among all existing identity-based encryption (IBE) schemes in bilinear groups, Wat-IBE proposed by Waters (CRYPTO 2009) and JR-IBE proposed by Jutla and Roy (Asiacrypt 2013) are quite special. A secret key and/or ciphertext in these two schemes consists of several group elements and an integer which is usually called tag.

A series of prior work was devoted to extending them towards more advanced attribute-based encryption (ABE) including inner-product encryption (IPE), hierarchical IBE (HIBE). Recently, Kim et al. (SCN 2016) introduced the notion of tag-based encoding and presented a generic framework for extending Wat-IBE. We may call these ABE schemes ABE with tag or tag-based ABE. Typically, a tag-based ABE construction is more efficient than its counterpart without tag. However, the research on tag-based ABE severely lags: we do not know how to extend JR-IBE in a systematic way and there is no tag-based ABE for Boolean span program even with Kim et al.'s generic framework.

This paper [32] proposes a generic framework for tag-based ABE which is based on JR-IBE and compatible with Chen et al.'s (attribute-hiding) predicate encoding (Eurocrypt 2015). The adaptive security in the standard model relies on the k-linear assumption in asymmetric prime-order bilinear groups. This is the first framework showing how to extend JR-IBE systematically. In fact, the framework and its simple extension are able to cover most concrete tag-based ABE constructions in previous literature. Furthermore, since Chen et al.'s predicate encoding supports a large number of predicates including boolean span program, the paper can give the first (both key-policy and ciphertext-policy) tag-based ABE for boolean span program in the standard model. Technically, the new framework is based on a simplified version of JR-IBE. Both the description and its proof are quite similar to the prime-order IBE derived from Chen et al.'s framework. This not only allows us to work with Chen et al.'s predicate encoding but also provides a clear explanation of the JR-IBE scheme and its proof technique.

7.2.8. Hardness of k-LWE and Applications in Traitor Tracing

The paper introduces the k-LWE problem, a Learning With Errors variant of the k-SIS problem. The Boneh-Freeman reduction from SIS to k-SIS suffers from an exponential loss in k. The paper [24] improves and extend it to an LWE to k-LWE reduction with a polynomial loss in k, by relying on a new technique involving trapdoors for random integer kernel lattices. Based on this hardness result, the paper presents the first algebraic construction of a traitor tracing scheme whose security relies on the worstcase hardness of standard lattice problems. The proposed LWE traitor tracing is almost as efficient as the LWE encryption. Further, it achieves public traceability, i.e., allows the authority to delegate the tracing capability to trusted parties. To this aim, the paper introduces the notion of projective sampling family in which each sampling function is keyed and, with a projection of the key on a well chosen space, one can simulate the sampling function in a computationally indistinguishable way. The construction of a projective sampling family from k-LWE allows us to achieve public traceability, by publishing the projected keys of the users.

7.2.9. Middle-Product Learning With Errors

The paper [45] introduces a new variant MPLWE of the Learning With Errors problem (LWE) making use of the Middle Product between polynomials modulo an integer q. It exhibits a reduction from the Polynomial-LWE problem (PLWE) parametrized by a polynomial f, to MPLWE which is defined independently of any such f. The reduction only requires f to be monic with constant coefficient coprime with q. It incurs a noise growth proportional to the so-called expansion factor of f. The paper also describes a public-key encryption scheme with quasi-optimal asymptotic efficiency (the bit-sizes of the keys and the run-times of all involved algorithms are quasi-linear in the security parameter), which is secure against chosen plaintext attacks under the MPLWE hardness assumption. The scheme is hence secure under the assumption that PLWE is hard for at least one polynomial f of degree n among a family of f's which is exponential in n.

7.2.10. Efficient Public Trace and Revoke from Standard Assumptions

The paper [27] provides efficient constructions for trace-and-revoke systems with public traceability in the black-box confirmation model. The constructions achieve adaptive security, are based on standard assumptions and achieve significant efficiency gains compared to previous constructions. The constructions rely on a generic transformation from inner product functional encryption (IPFE) schemes to trace-and-revoke systems. The proposed transformation requires the underlying IPFE scheme to only satisfy a very weak notion of security the attacker may only request a bounded number of random keys in contrast to the standard notion of security where she may request an unbounded number of arbitrarily chosen keys. The paper exploits the

much weaker security model to provide a new construction for bounded collusion and random key IPFE from the learning with errors assumption (LWE), which enjoys improved efficiency compared to the scheme of Agrawal et al. [CRYPTO'16]. Together with IPFE schemes from Agrawal et al., the paper obtains trace and revoke from LWE, Decision Diffie Hellman and Decision Quadratic Residuosity.

7.2.11. New Techniques for Structural Batch Verification in Bilinear Groups with Applications to Groth-Sahai Proofs

Bilinear groups form the algebraic setting for a multitude of important cryptographic protocols including anonymous credentials, e-cash, e-voting, e-coupon, and loyalty systems. It is typical of such crypto protocols that participating parties need to repeatedly verify that certain equations over bilinear groups are satisfied, e.g., to check that computed signatures are valid, commitments can be opened, or non-interactive zero-knowledge proofs verify correctly. Depending on the form and number of equations this part can quickly become a performance bottleneck due to the costly evaluation of the bilinear map.

To ease this burden on the verifier, batch verification techniques have been proposed that allow to combine and check multiple equations probabilistically using less operations than checking each equation individually. The paper [34] revisits the batch verification problem and existing standard techniques. It introduces a new technique which, in contrast to previous work, allows to fully exploit the structure of certain systems of equations. Equations of the appropriate form naturally appear in many protocols, e.g., due to the use of Groth–Sahai proofs.

The beauty of the new technique is that the underlying idea is pretty simple: the paper observes that many systems of equations can alternatively be viewed as a single equation of products of polynomials for which probabilistic polynomial identity testing following Schwartz–Zippel can be applied. Comparisons show that the new approach can lead to significant improvements in terms of the number of pairing evaluations. Indeed, for the BeleniosRF voting system presented at CCS 2016, it is possible to reduce the number of pairings (required for ballot verification) from $4k + 140$, as originally reported by Chaidos et al., to $k + 7$. As the implementation and benchmarks demonstrate, this may reduce the verification runtime to only 5% to 13% of the original runtime.

7.2.12. Encryption Switching Protocols Revisited: Switching Modulo p

At CRYPTO 2016, Couteau, Peters and Pointcheval introduced a new primitive called Encryption Switching Protocols (ESP), allowing to switch ciphertexts between two encryption schemes. If such an ESP is built with two schemes that are respectively additively and multiplicatively homomorphic, it naturally gives rise to a secure 2-party computation protocol. It is thus perfectly suited for evaluating functions, such as multivariate polynomials, given as arithmetic circuits. Couteau et al. built an ESP to switch between Elgamal and Paillier encryptions which do not naturally fit well together. Consequently, they had to design a clever variant of Elgamal over $\mathbf{Z}/n\mathbf{Z}$ with a costly shared decryption.

In [31], we first present a conceptually simple generic construction for encryption switching protocols. We then give an efficient instantiation of our generic approach that uses two well-suited protocols, namely a variant of Elgamal in $\mathbf{Z}/p\mathbf{Z}$ and the Castagnos-Laguillaumie encryption which is additively homomorphic over $\mathbf{Z}/p\mathbf{Z}$. Among other advantages, this allows to perform all computations modulo a prime p instead of an RSA modulus. Overall, our solution leads to significant reductions in the number of rounds as well as the number of bits exchanged by the parties during the interactive protocols. We also show how to extend its security to the malicious setting.

7.3. Algebraic computing and high-performance kernels

7.3.1. Multiple binomial sums

Multiple binomial sums form a large class of multi-indexed sequences, closed under partial summation, which contains most of the sequences obtained by multiple summation of binomial coefficients and also all the sequences with algebraic generating function. We study the representation of the generating functions of

binomial sums by integrals of rational functions. The outcome is twofold. Firstly, we show that a univariate sequence is a multiple binomial sum if and only if its generating function is the diagonal of a rational function. Secondly we propose algorithms that decide the equality of multiple binomial sums and that compute recurrence relations for them. In conjunction with geometric simplifications of the integral representations, this approach behaves well in practice. The process avoids the computation of certificates and the problem of accurate summation that afflicts discrete creative telescoping, both in theory and in practice [12].

7.3.2. Algebraic Diagonals and Walks: Algorithms, Bounds, Complexity

The diagonal of a multivariate power series F is the univariate power series $\mathrm{Diag}(F)$ generated by the diagonal terms of F. Diagonals form an important class of power series; they occur frequently in number theory, theoretical physics and enumerative combinatorics. We study algorithmic questions related to diagonals in the case where F is the Taylor expansion of a bivariate rational function. It is classical that in this case $\mathrm{Diag}(F)$ is an algebraic function. We propose an algorithm that computes an annihilating polynomial for $\mathrm{Diag}(F)$. We give a precise bound on the size of this polynomial and show that generically, this polynomial is the minimal polynomial and that its size reaches the bound. The algorithm runs in time quasi-linear in this bound, which grows exponentially with the degree of the input rational function. We then address the related problem of enumerating directed lattice walks. The insight given by our study leads to a new method for expanding the generating power series of bridges, excursions and meanders. We show that their first N terms can be computed in quasi-linear complexity in N, without first computing a very large polynomial equation [10].

7.3.3. Computing minimal interpolation bases

In [20] we consider the problem of computing univariate polynomial matrices over a field that represent minimal solution bases for a general interpolation problem, some forms of which are the vector M-Padé approximation problem in [Van Barel and Bultheel, Numerical Algorithms 3, 1992] and the rational interpolation problem in [Beckermann and Labahn, SIAM J. Matrix Anal. Appl. 22, 2000]. Particular instances of this problem include the bivariate interpolation steps of Guruswami-Sudan hard-decision and Kötter-Vardy soft-decision decodings of Reed-Solomon codes, the multivariate interpolation step of list-decoding of folded Reed-Solomon codes, and Hermite-Padé approximation. In the mentioned references, the problem is solved using iterative algorithms based on recurrence relations. Here, we discuss a fast, divide-and-conquer version of this recurrence, taking advantage of fast matrix computations over the scalars and over the polynomials. This new algorithm is deterministic, and for computing shifted minimal bases of relations between m vectors of size σ it uses $\widetilde{O}(m^{\omega-1}(\sigma+|s|))$ field operations, where ω is the exponent of matrix multiplication, $|s|$ is the sum of the entries of the input shift s with $\min(s)=0$, and the soft-O notation indicates that logarithmic factors in the big-O are omitted. This complexity bound improves in particular on earlier algorithms in the case of bivariate interpolation for soft decoding, while matching fastest existing algorithms for simultaneous Hermite-Padé approximation.

7.3.4. Fast and deterministic computation of the Hermite normal form and determinant of a polynomial matrix

Given a nonsingular $n \times n$ matrix of univariate polynomials over a field, we present in [22] fast and deterministic algorithms to compute its determinant and its Hermite normal form. The proposed algorithms use $\widetilde{O}(n^{\omega}\lceil s\rceil)$ field operations, where s is bounded from above by both the average of the degrees of the rows and that of the columns of the matrix, and ω is the exponent of matrix multiplication. The ceiling function indicates that the cost is $\widetilde{O}(n^{\omega})$ when $s=o(1)$. Our algorithms are based on a fast and deterministic triangularization method for computing the diagonal entries of the Hermite form of a nonsingular matrix.

7.3.5. Computing canonical bases of modules of univariate relations

We study in [44] the computation of canonical bases of sets of univariate relations $(p_1,...,p_m) \in K[x]^m$ such that $p_1 f_1 + \cdots + p_m f_m = 0$; here, the input elements $f_1,...,f_m$ are from a quotient $K[x]^n/\mathcal{M}$, where $\mathcal{M}$ is a $K[x]$-module of rank n given by a basis $M \in K[x]^{n\times n}$ in Hermite form. We exploit the triangular shape of M to generalize a divide-and-conquer approach which originates from fast minimal approximant basis algorithms. Besides recent techniques for this approach, we rely on high-order lifting to perform fast modular

products of polynomial matrices of the form $PF \bmod M$. Our algorithm uses $\widetilde{O}(m^{\omega-1}D + n^{\omega}D/m)$ operations in K, where $D = \deg(\det(M))$ is the K-vector space dimension of $K[x]^n/\mathcal{M}$, $\widetilde{O}(\cdot)$ indicates that logarithmic factors are omitted, and ω is the exponent of matrix multiplication. This had previously only been achieved for a diagonal matrix M. Furthermore, our algorithm can be used to compute the shifted Popov form of a nonsingular matrix within the same cost bound, up to logarithmic factors, as the previously fastest known algorithm, which is randomized.

7.3.6. Matrices with displacement structure: generalized operators and faster algorithms

For matrices with displacement structure, basic operations like multiplication, inversion, and linear system solving can be expressed in terms of the following task: evaluate the product AB, where A is a structured $n \times n$ matrix of displacement rank α, and B is an arbitrary $n \times \alpha$ matrix. In [11], we first generalize classical displacement operators, based on block diagonal matrices with companion diagonal blocks, and then design fast algorithms to perform the task above for this extended class of structured matrices. The arithmetic cost of these algorithms ranges from $O(\alpha^{\omega-1}M(n))$ to $O(\alpha^{\omega-1}M(n)\log(n))$, with ω such that two $n \times n$ matrices over a field can be multiplied using $O(n^{\omega})$ field operations, and where M is a cost function for polynomial multiplication. By combining this result with classical randomized regularization techniques, we obtain faster Las Vegas algorithms for structured inversion and linear system solving.

7.3.7. Absolute real root separation

While the separation (the minimal nonzero distance) between roots of a polynomial is a classical topic, its absolute counterpart (the minimal nonzero distance between their absolute values) does not seem to have been studied much. We present the general context and give tight bounds for the case of real roots [14].

7.3.8. Weighted Lattice Walks and Universality Classes

In this work we consider two different aspects of weighted walks in cones. To begin we examine a particular weighted model, known as the Gouyou-Beauchamps model. Using the theory of analytic combinatorics in several variables we obtain the asymptotic expansion of the total number of Gouyou-Beauchamps walks confined to the quarter plane. Our formulas are parametrized by weights and starting point, and we identify six different asymptotic regimes (called universality classes) which arise according to the values of the weights. The weights allowed in this model satisfy natural algebraic identities permitting an expression of the weighted generating function in terms of the generating function of unweighted walks on the same steps. The second part of this article explains these identities combinatorially for walks in arbitrary cones and dimensions, and provides a characterization of universality classes for general weighted walks. Furthermore, we describe an infinite set of models with non-D-finite generating function [15].

7.3.9. Introduction to the IEEE 1788-2015 Standard for Interval Arithmetic

Interval arithmetic is a tool of choice for numerical software verification, as every result computed using this arithmetic is self-verified: every result is an interval that is guaranteed to contain the exact numerical values, regardless of uncertainty or roundoff errors. From 2008 to 2015, interval arithmetic underwent a standardization effort, resulting in the IEEE 1788-2015 standard. The main features of this standard are developed in [26]: the structure into levels, from the mathematic model to the implementation on computers; the possibility to accomodate different mathematical models, called flavors; the decoration system that keeps track of relevant events during the course of a calculation; the exact dot product for point (as opposed to interval) vectors.

7.3.10. Influence of the Condition Number on Interval Computations: Some Examples

The condition number is a quantity that is well-known in "classical" numerical analysis, that is, where numerical computations are performed using floating-point numbers. This quantity appears much less frequently in interval numerical analysis, that is, where the computations are performed on intervals. In [56], two aspects are developed. On the one hand, it is stressed that the notion of condition number already appears in the literature on interval analysis, even if it does not bear that name. On the other hand, three small examples are used to illustrate experimentally the impact of the condition number on interval computations. As expected, problems

with a larger condition number are more difficult to solve: this means either that the solution is not very accurate (for moderate condition numbers) or that the method fails to solve the problem, even inaccurately (for larger condition numbers). Different strategies to counteract the impact of the condition number are discussed and experimented: use of a higher precision, iterative refinement, bisection of the input.

7.3.11. Error bounds on complex floating-point multiplication with an FMA

The accuracy analysis of complex floating-point multiplication done by Brent, Percival, and Zimmermann is extended to the case where a fused multiply-add (FMA) operation is available. Considering floating-point arithmetic with rounding to nearest and unit roundoff u, we show that their bound $\sqrt{5}\,u$ on the normwise relative error $|\widehat{z}/z - 1|$ of a complex product z can be decreased further to $2u$ when using the FMA in the most naive way. Furthermore, we prove that the term $2u$ is asymptotically optimal not only for this naive FMA-based algorithm, but also for two other algorithms, which use the FMA operation as an efficient way of implementing rounding error compensation. Thus, although highly accurate in the componentwise sense, these two compensated algorithms bring no improvement to the normwise accuracy $2u$ already achieved using the FMA naively. Asymptotic optimality is established for each algorithm thanks to the explicit construction of floating-point inputs for which we prove that the normwise relative error then generated satisfies $|\widehat{z}/z - 1| \to 2u$ as $u \to 0$. All our results hold for IEEE floating-point arithmetic, with radix β, precision p, and rounding to nearest; it is only assumed that underflows and overflows do not occur and that $\beta^{p-1} \geq 24$ [19].

7.3.12. Automatic source-to-source error compensation of floating-point programs

Numerical programs with IEEE 754 floating-point computations may suffer from inaccuracies, since finite precision arithmetic is an approximation of real arithmetic. Solutions that reduce the loss of accuracy are available, such as, compensated algorithms or double-double precision floating-point arithmetic. Our goal is to automatically improve the numerical quality of a numerical program with the smallest impact on its performance. In [25] we define and implement source code transformations in order to derive automatically compensated programs. We present several experimental results to compare the transformed programs and existing solutions. The transformed programs are as accurate and efficient as the implementations of compensated algorithms when the latter exist. Furthermore, we propose some transformation strategies allowing us to improve partially the accuracy of programs and to tune the impact on execution time. Trade-offs between accuracy and performance are assured by code synthesis. Experimental results show that, with the help of the tools presented here, user-defined trade-offs are achievable in a reasonable amount of time.

7.3.13. Formal correctness of comparison algorithms between binary64 and decimal64 floating-point numbers

We present a full Coq formalisation of the correctness of some comparison algorithms between binary64 and decimal64 floating-point numbers [28].

7.3.14. Implementation and performance evaluation of an extended precision floating-point arithmetic library for high-accuracy semidefinite programming

Semidefinite programming (SDP) is widely used in optimization problems with many applications, however, certain SDP instances are ill-posed and need more precision than the standard double-precision available. Moreover, these problems are large-scale and could benefit from parallelization on specialized architectures such as GPUs. In this article, we implement and evaluate the performance of a floating-point expansion-based arithmetic library (newFPLib) in the context of such numerically highly accurate SDP solvers. We plugged-in the newFPLib with the state-of-the-art SDPA solver for both CPU and GPU-tuned implementations. We compare and contrast both the numerical accuracy and performance of SDPA-GMP,-QD and-DD, which employ other multiple-precision arithmetic libraries against SDPA-newFPLib. We show that our newFPLib is a very good trade-off for accuracy and speed when solving ill-conditioned SDP problems [38].

7.3.15. The classical relative error bounds for computing $\sqrt{a^2+b^2}$ and $c/\sqrt{a^2+b^2}$ in binary floating-point arithmetic are asymptotically optimal

We study the accuracy of classical algorithms for evaluating expressions of the form $\sqrt{a^2+b^2}$ and $c/\sqrt{a^2+b^2}$ in radix-2, precision-p floating-point arithmetic, assuming that the elementary arithmetic operations $\pm$, $\times$, $/$, $\sqrt{}$ are rounded to nearest, and assuming an unbounded exponent range. Classical analyses show that the relative error is bounded by $2u+\mathcal{O}(u^2)$ for $\sqrt{a^2+b^2}$, and by $3u+\mathcal{O}(u^2)$ for $c/\sqrt{a^2+b^2}$, where $u=2^{-p}$ is the unit roundoff. Recently, it was observed that for $\sqrt{a^2+b^2}$ the $\mathcal{O}(u^2)$ term is in fact not needed. We show here that it is not needed either for $c/\sqrt{a^2+b^2}$. Furthermore, we show that these error bounds are asymptotically optimal. Finally, we show that the possible availability of an FMA instruction does not change the bounds, nor their asymptotic optimality [37].

7.3.16. On the relative error of computing complex square roots in floating-point arithmetic

We study the accuracy of a classical approach to computing complex square-roots in floating-point arithmetic. Our analyses are done in binary floating-point arithmetic in precision p, and we assume that the (real) arithmetic operations $+$, $-$, $\times$, $\div$, $\sqrt{}$ are rounded to nearest, so the unit roundoff is $u=2^{-p}$. We show that in the absence of underflow and overflow, the componentwise and normwise relative errors of this approach are at most $\frac{7}{2}u$ and $\frac{\sqrt{37}}{2}u$, respectively, and this without having to neglect terms of higher order in u. We then provide some input examples showing that these bounds are reasonably sharp for the three basic binary interchange formats (binary32, binary64, and binary128) of the IEEE 754 standard for floating-point arithmetic.

7.3.17. More accurate complex multiplication for embedded processors

In [36] we present some work in progress on the development of fast and accurate support for complex floating-point arithmetic on embedded processors. Focusing on the case of multiplication, we describe algorithms and implementations for computing both the real and imaginary parts with high relative accuracy. We show that, in practice, such accuracy guarantees can be achieved with reasonable overhead compared with conventional algorithms (which are those offered by current implementations and for which the real or imaginary part of a product can have no correct digit at all). For example, the average execution-time overheads when computing an FFT on the ARM Cortex-A53 and -A57 processors range from 1.04x to 1.17x only, while arithmetic costs suggest overheads from 1.5x to 1.8x.

7.3.18. Tight and rigorous error bounds for basic building blocks of double-word arithmetic

We analyze several classical basic building blocks of double-word arithmetic (frequently called "double-double arithmetic" in the literature): the addition of a double-word number and a floating-point number, the addition of two double-word numbers, the multiplication of a double-word number by a floating-point number, the multiplication of two double-word numbers, the division of a double-word number by a floating-point number, and the division of two double-word numbers. For multiplication and division we get better relative error bounds than the ones previously published. For addition of two double-word numbers, we show that the previously published bound was incorrect, and we provide a new relative error bound. We introduce new algorithms for division. We also give examples that illustrate the tightness of our bounds [21].

7.3.19. On the robustness of the 2Sum and Fast2Sum algorithms

The 2Sum and Fast2Sum algorithms are important building blocks in numerical computing. They are used (implicitely or explicitely) in many compensated algorithms (such as compensated summation or compensated polynomial evaluation). They are also used for manipulating floating-point expansions. We show that these algorithms are much more robust than it is usually believed: The returned result makes sense even when the rounding function is not round-to-nearest, and they are almost immune to overflow [9].

7.3.20. Formal verification of a floating-point expansion renormalization algorithm

Many numerical problems require a higher computing precision than the one offered by standard floating-point formats. A common way of extending the precision is to use floating-point expansions. As the problems may be critical and as the algorithms used have very complex proofs (many sub-cases), a formal guarantee

of correctness is a wish that can now be fulfilled, using interactive theorem proving. In this article we give a formal proof in Coq for one of the algorithms used as a basic brick when computing with floating-point expansions, the renormalization, which is usually applied after each operation. It is a critical step needed to ensure that the resulted expansion has the same property as the input one, and is more "compressed". The formal proof uncovered several gaps in the pen-and-paper proof and gives the algorithm a very high level of guarantee [30].

7.3.21. Interactive proof protocols

We present in [46] an interactive probabilistic proof protocol that certifies in $(\log N)^{O(1)}$ arithmetic and Boolean operations for the verifier for example the determinant of an $N \times N$ matrix over a field whose entries are given by a single $(\log N)^{O(1)}$-depth arithmetic circuit, which contains $(\log N)^{O(1)}$ field constants and which is polynomial time uniform. The prover can produce the interactive certificate within a $(\log N)^{O(1)}$ factor of the cost of computing the determinant. Our protocol is a version of the proofs for muggles protocol by Goldwasser, Kalai and Rothblum [STOC 2008, J. ACM 2015]. More generally, our verifier checks a computation on a family of circuits of size $N^{O(1)}$, or even $2^{(\log N)^{O(1)}}$, for $g_N(f_N(0), ..., f_N(N-1))$ in $(\log N)^{O(1)}$ bit communication and bit-operation complexity. Here g_N is a family of $(\log N)^{O(1)}$-depth circuits, and f_N is a family of $(\log N)^{O(1)}$-depth circuits for the scalars (such as hypergeometric terms); f_N can contain $(\log N)^{O(1)}$ input field constants. If the circuits f_N for the scalars are of size $(\log N)^{O(1)}$, they are input for the verifier. The circuit g_N and in the general case f_N are $N^{O(1)}$-sized and cannot be built by the verifier with poly-log complexity. The verifier rather accesses the circuits via algorithms that probe the circuit structures, which are called uniformity properties.

7.3.22. New development on GNU MPFR

Work on the new fast, low-level algorithm to compute the correctly rounded summation of several floating-point numbers in arbitrary precision in radix 2 (each number having its own precision), and its implementation in GNU MPFR (new `mpfr_sum` function), has been completed [23].

The basic operations of GNU MPFR have also been optimized in small precision, and faithful rounding (mainly for internal use) is now partly supported [39].

These improvements, among many other ones, will be available in GNU MPFR 4.0.0; a release candidate is distributed in December 2017.

AROMATH Project-Team

6. New Results

6.1. Waring-like decompositions of polynomials

Participant: Alessandro Oneto.

In [9], we consider particular types of *additive decompositions* of homogeneous polynomials. The classical decomposition is the *Waring decomposition*, where we decompose polynomials as sums of powers of linear forms. Another well studied decomposition is the sometimes-called *Chow decomposition*, where we decompose polynomials as sums of products of linear forms. These are the extremal cases of the additive decompositions considered in this work. For a fixed partition $(d_1, ..., d_s) \vdash d$ of the degree of the polynomial, we consider decompositions as sums of degree forms of the form $\ell_1^{d_1} \cdots \ell_s^{d_s}$, where the ℓ's are linear forms. The homogeneous polynomials of the form $\ell_1^{d_1} \cdots \ell_s^{d_s}$ are parametrized by particular linear projections of certain Segre-Veronese varieties. The main results of this work concerns the dimension of the secant varieties to these projections of Segre-Veronese varieties. In particular, we compute their dimensions in the binary case (forms in two variables) and the case of secant lines varieties for any partition and any number of variables. From these results, we deduce the dimension of higher secant varieties in some particular cases.

This is a joint work with M. V. Catalisano, Luca Chiantini, and A. V. Geramita.

6.2. Waring loci and the Strassen conjecture

Participant: Alessandro Oneto.

In [8], we introduce the notion of the *Waring locus* of a homogeneous polynomial. A *Waring decomposition* is an expression of a polynomial as sum of powers of linear forms. The smallest length of such a decomposition is called the *Waring rank* of the polynomial. A very difficult challenge is to compute the rank and a minimal decomposition of a given form. The Waring locus of a polynomial is the locus of linear forms that appear in a minimal decomposition of it. The idea behind this construction is to find an iterative approach to construct Waring decompositions *step-by-step*, by adding one power at the time. Moreover, we give a version of the famous *Strassen conjecture* on the additivity of rank for sums of polynomials in independent sets of variables. We compute the Waring loci in several cases as binary forms, quadrics, monomials and plane cubics and for some other particular families of polynomials.

This is a joint work with E. Carlini, and M. V. Catalisano.

6.3. Minkowski sums and Hadamard products of algebraic varieties

Participant: Alessandro Oneto.

In [26], we study two particular geometric constructions. Given two affine algebraic varieties, we define their *Minkowski sum* as the (Zariski) closure of the set of coefficient-wise sums of pairs of points in the two varieties. Given two projective varieties, we define their *Hadamard product* as the (Zariski) closure of the set of coefficient-wise multiplications of pairs of points in the two varieties. In particular, we focus on computing their dimensions and degrees in terms of the ones of the original varieties. Hadamard products are of particular interests as they can be used to parametrize particular families of tensors which rise naturally by studying Restricted Boltzmann Machines, which are particular structures used in Statistics and Machine Learning.

This is a joint work with N. Friedenberg, and R. Williams.

6.4. Polynomial-exponential decomposition from moments

Participant: Bernard Mourrain.

In [12], we analyze the decomposition problem of multivariate polynomial-exponential functions from truncated series and present new algorithms to compute their decomposition. Using the duality between polynomials and formal power series, we first show how the elements in the dual of an Artinian algebra correspond to polynomial-exponential functions. They are also the solutions of systems of partial differential equations with constant coefficients. We relate their representation to the inverse system of the roots of the characteristic variety. Using the properties of Hankel operators, we establish a correspondence between polynomial exponential series and Artinian Gorenstein algebras. We generalize Kronecker theorem to the multivariate case, by showing that the symbol of a Hankel operator of finite rank is a polynomial-exponential series and by connecting the rank of the Hankel operator with the decomposition of the symbol. A generalization of Prony's approach to multivariate decomposition problems is presented, exploiting eigenvector methods for solving polynomial equations. We show how to compute the frequencies and weights of a minimal polynomial-exponential decomposition, using the first coefficients of the series. A key ingredient of the approach is the flat extension criteria, which leads to a multivariate generalization of a rank condition for a Carathéodory-Fejér decomposition of multivariate Hankel matrices. A new algorithm is given to compute a basis of the Artinian Gorenstein algebra, based on a Gram-Schmidt orthogonalization process and to decompose polynomial-exponential series. A general framework for the applications of this approach is described and illustrated in different problems. We provide Kronecker-type theorems for convolution operators, showing that a convolution operator (or a cross-correlation operator) is of finite rank, if and only if, its symbol is a polynomial-exponential function, and we relate its rank to the decomposition of its symbol. We also present Kronecker-type theorems for the reconstruction of measures as weighted sums of Dirac measures from moments and for the decomposition of polynomial-exponential functions from values. Finally, we describe an application of this method for the sparse interpolation of polylog functions from values.

6.5. Fast algorithm for border bases of Artinian Gorenstein algebras

Participant: Bernard Mourrain.

Given a multi-index sequence σ, we present in [23] a new efficient algorithm to compute generators of the linear recurrence relations between the terms of σ. We transform this problem into an algebraic one, by identifying multi-index sequences, multivariate formal power series and linear functionals on the ring of multivariate polynomials. In this setting, the recurrence relations are the elements of the kernel I_σ of the Hankel operator H_σ associated to σ. We describe the correspondence between multi-index sequences with a Hankel operator of finite rank and Artinian Gorenstein Algebras. We show how the algebraic structure of the Artinian Gorenstein algebra A_σ associated to the sequence σ yields the structure of the terms σ_α for all $\alpha \in \mathbb{N}^n$. This structure is explicitly given by a border basis of A_σ, which is presented as a quotient of the polynomial ring $K[x_1, ..., x_n]$ by the kernel I_σ of the Hankel operator H_σ. The algorithm provides generators of I_σ constituting a border basis, pairwise orthogonal bases of A_σ and the tables of multiplication by the variables in these bases. It is an extension of Berlekamp-Massey-Sakata (BMS) algorithm, with improved complexity bounds. We present applications of the method to different problems such as the decomposition of functions into weighted sums of exponential functions, sparse interpolation, fast decoding of algebraic codes, computing the vanishing ideal of points, and tensor decomposition. Some benchmarks illustrate the practical behavior of the algorithm.

6.6. Structured low rank decomposition of multivariate Hankel matrices

Participants: Jouhayna Harmouch, Bernard Mourrain.

In [11], we study the decomposition of a multivariate Hankel matrix H_σ as a sum of Hankel matrices of small rank in correlation with the decomposition of its symbol σ as a sum of polynomial-exponential series. We present a new algorithm to compute the low rank decomposition of the Hankel operator and the decomposition of its symbol exploiting the properties of the associated Artinian Gorenstein quotient algebra A_σ. A basis of A_σ is computed from the Singular Value Decomposition of a sub-matrix of the Hankel matrix H_σ. The frequencies and the weights are deduced from the generalized eigenvectors of pencils of shifted sub-matrices of H_σ. Explicit formula for the weights in terms of the eigenvectors avoid us to solve a Vandermonde system. This new

method is a multivariate generalization of the so-called Pencil method for solving Prony-type decomposition problems. We analyze its numerical behavior in the presence of noisy input moments, and describe a rescaling technique which improves the numerical quality of the reconstruction for frequencies of high amplitudes. We also present a new Newton iteration, which converges locally to the closest multivariate Hankel matrix of low rank and show its impact for correcting errors on input moments.

This is a joint work with Houssam Khalil.

6.7. Decomposition of low rank multi-symmetric tensor

Participants: Jouhayna Harmouch, Bernard Mourrain.

In [22], we study the decomposition of a multi-symmetric tensor T as a sum of powers of product of linear forms in correlation with the decomposition of its dual T^* as a weighted sum of evaluations. We use the properties of the associated Artinian Gorenstein Algebra A_τ to compute the decomposition of its dual T^* which is defined via a formal power series τ. We use the low rank decomposition of the Hankel operator $\mathrm{H}\tau$ associated to the symbol τ into a sum of indecomposable operators of low rank. A basis of A_τ is chosen such that the multiplication by some variables is possible. We compute the sub-coordinates of the evaluation points and their weights using the eigen-structure of multiplication matrices. The new algorithm that we propose works for small rank. We give a theoretical generalized approach of the method in n dimensional space. We show a numerical example of the decomposition of a multi-linear tensor of rank 3 in 3 dimensional space.

This is a joint work with Houssam Khalil.

6.8. Tensor decomposition and homotopy continuation

Participant: Bernard Mourrain.

A computationally challenging classical elimination theory problem is to compute polynomials which vanish on the set of tensors of a given rank. By moving away from computing polynomials via elimination theory to computing pseudowitness sets via numerical elimination theory, we develop in [3] computational methods for computing ranks and border ranks of tensors along with decompositions. More generally, we present our approach using joins of any collection of irreducible and nondegenerate projective varieties $X_1, ..., X_k \subset \mathbb{P}^N$ defined over $\mathbb{C}$. After computing ranks over $\mathbb{C}$, we also explore computing real ranks. Various examples are included to demonstrate this numerical algebraic geometric approach.

This is a joint work with Alessandra Bernardi, Noah S. Daleo, Jonathan D. Hauenstein.

6.9. Effective criteria for bigraded birational maps

Participant: Laurent Busé.

In [6], we consider rational maps whose source is a product of two subvarieties, each one being embedded in a projective space. Our main objective is to investigate birationality criteria for such maps. First, a general criterion is given in terms of the rank of a couple of matrices that became to be known as Jacobian dual matrices. Then, we focus on rational maps from the product of two projective lines to the projective plane in very low bidegrees and provide new matrix-based birationality criteria by analyzing the syzygies of the defining equations of the map, in particular by looking at the dimension of certain bigraded parts of the syzygy module. Finally, applications of our results to the context of geometric modeling are discussed at the end of the paper.

This is a joint work with Nicolás Botbol (University of Buenos Aires), Marc Chardin (University of Paris VI), Hamid Seyed Hassanzadeh (University of Rio de Janeiro), Aron Simis (University of Pernambuci) and Quang Hoa Tran (University of Paris VI).

6.10. Discriminants of complete intersection space curves

Participant: Laurent Busé.

In [19], we develop a new approach to the discriminant of a complete intersection curve in the 3-dimensional projective space. By relying on the resultant theory, we first prove a new formula that allows us to define this discriminant without ambiguity and over any commutative ring, in particular in any characteristic. This formula also provides a new method for evaluating and computing this discriminant efficiently, without the need to introduce new variables as with the well-known Cayley trick. Then, we obtain new properties and computational rules such as the covariance and the invariance formulas. Finally, we show that our definition of the discriminant satisfies to the expected geometric property and hence yields an effective smoothness criterion for complete intersection space curves. Actually, we show that in the generic setting, it is the defining equation of the discriminant scheme if the ground ring is assumed to be a unique factorization domain.

This is a joint work with Ibrahim Nonkané (University of Ouaga II).

6.11. Matrix Representations by Means of Interpolation

Participants: Ioannis Emiris, Christos Konaxis, Clément Laroche.

In [20] we examine implicit representations of parametric or point cloud models, based on interpolation matrices, which are not sensitive to base points. We show how interpolation matrices can be used for ray shooting of a parametric ray with a surface patch, including the case of high-multiplicity intersections. Most matrix operations are executed during pre-processing since they solely depend on the surface. For a given ray, the bottleneck is equation solving. Our Maple code handles bicubic patches within 1 second, though numerical issues might arise. Our second contribution is to extend the method to parametric space curves and, generally, to codimension > 1, by computing the equations of (hyper)surfaces intersecting precisely at the given object. By means of Chow forms, we propose a new, practical, randomized algorithm that always produces correct output but possibly with a non-minimal number of surfaces. For space curves, we obtain 3 surfaces whose polynomials are of near-optimal degree; in this case, computation reduces to a Sylvester resultant. We illustrate our algorithm through a series of examples and compare our Maple prototype with other methods implemented in Maple, i.e., Gröbner basis and implicit matrix representations. Our Maple prototype is not faster but yields fewer equations and seems more robust than Maple's implicitize; it is also comparable with the other methods for degrees up to 6.

Joint work with I.S. Kotsireas.

6.12. Efficient certication of numeric solutions to eigenproblems

Participant: Bernard Mourrain.

In [24], we present an efficient algorithm for the certication of numeric solutions to eigenproblems. The algorithm relies on a mixture of ball arithmetic, a suitable Newton iteration, and clustering of eigenvalues that are close.

This is a joint work with Joris Van Der Hoeven.

6.13. Approximating multidimensional subset sum and minkowski decomposition of polygons

Participants: Ioannis Emiris, Anna Karasoulou.

In [10] we consider the approximation of two NP-hard problems: Minkowski Decomposition (MinkDecomp) of lattice polygons in the plane and the closely related problem of Multidimensional Subset Sum (kD-SS) in arbitrary dimension. In kD-SS we are given an input set S of k-dimensional vectors, a target vector t and we ask if there exists a subset of S that sums up to t. We prove, through a gap-preserving reduction, that, for general dimension k, kD-SS is not in APX although the classic 1D-SS is in PTAS. On the positive side, we present an $O(n^3/\epsilon^2)$ approximation grid based algorithm for 2D-SS, where n is the cardinality of the set and $\epsilon > 0$ bounds the difference of some measure of the input polygon and the sum of the output polygons. We also describe two approximation algorithms with a better experimental ratio. Applying one of these algorithms, and

a transformation from MinkDecomp to 2D-SS, we can approximate Mink-Decomp. For an input polygon Q and parameter ϵ, we return two summands A and B such that $A + B = Q'$ with Q' being bounded in relation to Q in terms of volume, perimeter, or number of internal lattice points, an additive error linear in and up to quadratic in the diameter of Q. A similar function bounds the Hausdorff distance between Q and Q'. We offer experimental results based on our implementation.

Joint with C. Tzovas.

6.14. High-dimensional approximate r-nets

Participants: Ioannis Emiris, Ioannis Psarros.

The construction of r-nets offers a powerful tool in computational and metric geometry. In [17], we focus on high-dimensional spaces and present a new randomized algorithm which efficiently computes approximate r-nets with respect to Euclidean distance. For any fixed $\epsilon > 0$, the approximation factor is $1 + \epsilon$ and the complexity is polynomial in the dimension and subquadratic in the number of points. The algorithm succeeds with high probability. More specifically, the best previously known LSH-based construction is improved in terms of complexity by reducing the dependence on ϵ, provided that ϵ is sufficiently small. Our method does not require LSH but, instead, follows Valiant's (2015) approach in designing a sequence of reductions of our problem to other problems in different spaces, under Euclidean distance or inner product, for which r-nets are computed efficiently and the error can be controlled. Our result immediately implies efficient solutions to a number of geometric problems in high dimension, such as finding the $(1 + \epsilon)$-approximate kth nearest neighbor distance in time subquadratic in the size of the input.

Joint with G. Avarikioti, L. Kavouras.

6.15. Extraction of tori from minimal point sets

Participants: Laurent Busé, André Galligo.

In [7], a new algebraic method for extracting tori from a minimal point set, made of two oriented points and a simple point, is proposed. We prove a degree bound on the number of such tori; this bound is reached on examples, even when we restrict to smooth tori. Our method is based on pre-computed closed formulae well suited for numerical computations with approximate input data.

6.16. Scaffolding skeletons using spherical Voronoi diagrams

Participants: Alvaro Fuentes Suarez, Evelyne Hubert.

Given a skeleton made of line segments we describe how to obtain a coarse mesh (or scaffold) of a surface surrounding it. We emphasize in [21] the key result that allows us to complete a previous approach that could not treat skeletons with cycles.

6.17. G^1-smooth splines on quad meshes with 4-split macro-patch elements

Participants: Ahmed Blidia, Bernard Mourrain.

We analyze the space of differentiable functions on a quad-mesh $\mathcal{M}$, which are composed of 4-split spline macro-patch elements on each quadrangular face. We describe explicit transition maps across shared edges, that satisfy conditions which ensure that the space of differentiable functions is ample on a quad-mesh of arbitrary topology. These transition maps define a finite dimensional vector space of G^1 spline functions of bi-degree $\leq (k, k)$ on each quadrangular face of $\mathcal{M}$. We determine the dimension of this space of G^1 spline functions for k big enough and provide explicit constructions of basis functions attached respectively to vertices, edges and faces. This construction requires the analysis of the module of syzygies of univariate b-spline functions with b-spline function coefficients. New results on their generators and dimensions are provided. Examples of bases of G^1 splines of small degree for simple topological surfaces are detailed and illustrated by parametric surface constructions.

This is a joint work with Nelly Villamizar

6.18. Hermite type spline spaces over rectangular meshes with complex topological structures

Participants: André Galligo, Bernard Mourrain.

Motivated by the Magneto HydroDynamic (MHD) simulation for Tokamaks with Isogeometric analysis, we present in [14] a new type of splines defined over a rectangular mesh with arbitrary topology, which are piecewise polynomial functions of bidegree (d, d) and C^r parameter continuity. In particular, we compute their dimension and exhibit basis functions called Hermite bases for bicubic spline spaces. We investigate their potential applications for solving partial differential equations (PDEs) over a complex physical domain in the framework of Isogeometric analysis. In particular, we analyze the property of approximation of these spline spaces for the L^2-norm. Despite the fact that the basis functions are singular at extraordinary vertices, we show that the optimal approximation order and numerical convergence rates are reached by setting a proper parameterization.

This is a joint work with Meng Wu, Bernard Mourrain, André Galligo, Boniface Nkonga

6.19. H^1-parameterizations of complex planar physical domains in isogeometric analysis

Participants: André Galligo, Bernard Mourrain.

Isogeometric analysis (IGA) is a method for solving geometric partial differential equations (PDEs). Generating parameterizations of a PDE's physical domain is the basic and important issues within IGA framework. In [13], we present a global H^1-parameterization method for a planar physical domain with complex topology.

This is a joint work with Meng Wu, Boniface Nkonga.

6.20. Convergence rates with singular parameterizations for solving elliptic boundary value problems in isogeometric analysis

Participant: Bernard Mourrain.

In [15], we present convergence rates for solving elliptic boundary value problems with singular parameterizations in isogeometric analysis. First, the approximation errors with the $L^2(\Omega)$-norm and the $H^1(\Omega)$-seminorm are estimated locally. The impact of singularities is considered in this framework. Second, the convergence rates for solving PDEs with singular parameterizations are discussed. These results are based on a weak solution space that contains all of the weak solutions of elliptic boundary value problems with smooth coefficients. For the smooth weak solutions obtained by isogeometric analysis with singular parameterizations and the finite element method, both are shown to have the optimal convergence rates. For non-smooth weak solutions, the optimal convergence rates are reached by setting proper singularities of a controllable parameterization, even though convergence rates are not optimal by finite element method, and the convergence rates by isogeometric analysis with singular parameterizations are better than the ones by the finite element method.

This a joint work with Meng Wu, Yicao Wang, Boniface Nkonga, Changzheng Cheng.

6.21. Geometric modeling and deformation for shape optimization of ship hulls and appendages

Participants: Elisa Berrini, Bernard Mourrain.

The precise control of geometric models plays an important role in many domains such as computer-aided geometric design and numerical simulation. For shape optimization in computational fluid dynamics (CFD), the choice of control parameters and the way to deform a shape are critical. In [4], we describe a skeleton-based representation of shapes adapted for CFD simulation and automatic shape optimization. Instead of using the control points of a classic B-spline representation, we control the geometry in terms of architectural parameters. We assure valid shapes with a strong shape consistency control. Deformations of the geometry are performed by solving optimization problems on the skeleton. Finally, a surface reconstruction method is proposed to evaluate the shape's performances with CFD solvers. We illustrate the approach on two problems: the foil of an AC45 racing sail boat and the bulbous bow of a fishing trawler. For each case, we obtain a set of shape deformations and then we evaluate and analyzed the performances of the different shapes with CFD computations.

This is a joint work with Yann Roux, Matthieu Durand, Guillaume Fontaine.

6.22. Geometric model for automated multi-objective optimization of foils

Participants: Elisa Berrini, Bernard Mourrain.

The work in [18] describes a new generic parametric modeler integrated into an automated optimization loop for shape optimization. The modeler enables the generation of shapes by selecting a set of design parameters that controls a twofold parameterization: geometrical - based on a skeleton approach - and architectural - based on the experience of practitioners - to impact the system performance. The resulting forms are relevant and effective, thanks to a smoothing procedure that ensures the consistency of the shapes produced. As an application, we propose to perform a multi-objective shape optimization of an AC45 foil. The modeler is linked to the fluid solver AVANTI, coupled with Xfoil, and to the optimization toolbox FAMOSA.

This is a joint work with Régis Duvigneau, Matthieu Sacher, Yann Roux.

CARAMBA Project-Team

7. New Results

7.1. Improved Complexity Bounds for Counting Points on Hyperelliptic Curves

Participants: Simon Abelard, Pierrick Gaudry, Pierre-Jean Spaenlehauer.

In [16], we present a probabilistic Las Vegas algorithm for computing the local zeta function of a hyperelliptic curve of genus g defined over $\mathbb{F}_q$. It is based on the approaches by Schoof and Pila combined with a modeling of the ℓ-torsion by structured polynomial systems. Our main result improves on previously known complexity bounds by showing that there exists a constant $c > 0$ such that, for any fixed g, this algorithm has expected time and space complexity $O((\log q)^{cg})$ as q grows and the characteristic is large enough.

7.2. Deciphering of a Code Used by a 19th Century Parisian Violin Dealer

Participant: Pierrick Gaudry.

This paper [4] is joint work with Jean-Philippe Échard, Curator at the Cité de la Musique, Paris.

The study of three ledgers from the archives of a prominent Parisian violin maker's workshop (active from 1796 to 1948) reveals that some of their content was encrypted. We present the deciphering of the code, and a discussion of its use in the context of the workshop. Charles-Adolphe Gand introduced this code around 1847 to encrypt values of antique/used violins he would buy and resell. His successors maintained the use of this code at least until 1921. Taking a few examples of instruments by Stradivari and other violin makers, we illustrate how the decoded ledgers – listing transactions for more than 2,500 instruments – are of high interest as historical sources documenting the margins, rebates, and commercial practices of these violin dealers. More generally, we contribute to better describing the evolution of the market for antique instruments of the violin family.

7.3. Discrete Logarithm Record Computation in Extension Fields

Participants: Laurent Grémy, Aurore Guillevic, Emmanuel Thomé.

Together with F. Morain from the GRACE team, we reached new record sizes for the discrete logarithm problems over non-prime finite fields of small extension degrees [19], [8]. Assessing the hardness of the discrete logarithm problem in such fields is highly relevant to the security of cryptographic pairings. Our computations are not terribly large computations compared to other record-size computations for integer factoring or discrete logarithm over prime fields, but on the other hand more novelty is present in these contexts: use of automorphisms, higher degree sieving, for example.

Further research in this direction is needed, especially regarding the effectiveness of the variants of the "tower" number field sieve variants.

Furthermore, A. Guillevic and L. Grémy have gathered in a database all published records of discrete logarithm computations in all kinds of finite fields. The database is hosted on gitlab and is open to external contributions. A web interface for browsing the database is available at http://perso.ens-lyon.fr/laurent.gremy/dldb/index.html.

7.4. Using Constraint Programming to Solve a Cryptanalytic Problem

Participant: Marine Minier.

In [7], we describe Constraint Programming (CP) models to solve a cryptanalytic problem: the related key differential attack against the standard block cipher AES. We show that CP solvers are able to solve these problems quicker than dedicated cryptanalysis tools, and we prove that the 11 rounds solution on AES-192 claimed to be optimal is wrong. Instead, we provide the best related key differential characteristic on 10 rounds of AES-192. We also improved the related-key distinguisher and the basic related-key differential attack on the full AES-256 by a factor 2^6 and the q-multicollisions by a factor 2.

7.5. Optimized Binary64 and Binary128 Arithmetic with GNU MPFR

Participant: Paul Zimmermann.

Together with Vincent Lefèvre (ARIC team, Inria Rhône-Alpes), Paul Zimmermann wrote an article "Optimized Binary64 and Binary128 Arithmetic with GNU MPFR", and presented it at the 24th IEEE Symposium on Computer Arithmetic [9]. This article describes algorithms used to optimize the GNU MPFR library when the operands fit into one or two words. On modern processors, a correctly rounded addition of two quadruple precision numbers is now performed in 22 cycles, a subtraction in 24 cycles, a multiplication in 32 cycles, a division in 64 cycles, and a square root in 69 cycles. It also introduces a new faithful rounding mode, which enables even faster computations. These optimizations will be available in version 4 of MPFR.

7.6. A New Measure for Root Optimization

Participants: Nicolas David, Paul Zimmermann.

In the General Number Field Sieve (GNFS) for integer factorization or discrete logarithm, the first stage is polynomial selection. Polynomial selection itself consists in two steps: size-optimization and root-optimization. The classical measures used to rank polynomials during the root-optimization are the so-called α and Murphy-E values. During the internship of Nicolas David, it was shown that these classical measures might be off by up to 15% between two polynomial pairs, compared to a sieving test. A new measure that better corresponds to sieving tests was designed. An article describing these new results is in preparation.

7.7. Mathematical Computation with SageMath

Participant: Paul Zimmermann.

Starting in March, Paul Zimmermann coordinated the English translation of the book "Calcul mathématique avec Sage", and the update from version 5.9 to 8.0 of Sage. He also translated several chapters and proof-read the translation of all chapters. The current state of the English translation is available under a Creative Commons license (CC BY-SA) at https://members.loria.fr/PZimmermann/sagebook/english.html. A discussion is in process with an editor to publish a paper version.

7.8. Topics in Computational Number Theory Inspired by Peter L. Montgomery

Participants: Emmanuel Thomé, Paul Zimmermann.

Emmanuel Thomé and Paul Zimmermann contributed two chapters of the book "Topics in Computational Number Theory Inspired by Peter L. Montgomery", coordinated by Arjen Lenstra and Joppe Bos, and published by Cambridge University Press. Together with Richard P. Brent and Alexander Kruppa, Paul Zimmermann wrote a chapter entitled "FFT extension for algebraic-group factorization algorithms" [12]. Emmanuel Thomé contributed a chapter entitled "The block Lanczos algorithm" [14].

7.9. Improved Methods for Finding Optimal Formulae for Bilinear Maps in a Finite Field

Participant: Svyatoslav Covanov.

In [17], we describe a method improving on the exhaustive search algorithm developed in [26]. We are able to compute new optimal formulae for the short product modulo X^5 and the circulant product modulo $(X^5 - 1)$. Moreover, we prove that there is essentially only one optimal decomposition of the product of 3×2 by 2×3 matrices up to the action of some group of automorphisms.

7.10. Big Prime Field FFT on the GPU

Participant: Svyatoslav Covanov.

In colloboration with L. Chen, D. Mohajerani and M. Moreno Maza, in [11], we compare various methods for the multiplication of polynomials, using the GPU. We compare the CRT method, using k machine-word primes, to the generalized Fermat prime method, for a prime of k machine-words, inspired by the work in [28]. For some degrees and k, we prove that the arithmetic operations with the generalized Fermat primes offer attractive performance both in terms of algebraic complexity and parallelism.

7.11. CM Plane Quartics

Participant: Hugo Labrande.

As a by-product of his PhD thesis defended in late 2016, Hugo Labrande contributed to a joint work with several authors, leading to an article [21] that provides examples of smooth plane quartics over $\mathbb{Q}$ with complex multiplication over $\overline{\mathbb{Q}}$ by a maximal order with primitive CM type. Several algorithms are used, in tight connection to the computation of Theta functions which was improved in Labrande's PhD thesis: reduction of period matrices, fast computation of Dixmier-Ohno invariants, and reconstruction from these invariants.

7.12. Explicit Isogenies in Genus 2 and 3

Participant: Enea Milio.

In [22], we present a quasi-linear algorithm to compute isogenies between Jacobians of curves of genus 2 and 3 starting from the equation of the curve and a maximal isotropic subgroup of the ℓ-torsion, for ℓ an odd prime number, generalizing Vélu's formula of genus 1. This work is based on the paper "Computing functions on Jacobians and their quotients" of Jean-Marc Couveignes and Tony Ezome. We improve their genus 2 case algorithm, generalize it for genus 3 hyperelliptic curves and introduce a way to deal with the genus 3 non-hyperelliptic case, using algebraic Theta functions.

7.13. Modular Polynomials of Hilbert Surfaces

Participant: Enea Milio.

In [23], together with Damien Robert from the LFANT team, we describe an evaluation/interpolation approach to compute modular polynomials on a Hilbert surface, which parametrizes abelian surfaces with maximal real multiplication. Under some heuristics we obtain a quasi-linear algorithm. The corresponding modular polynomials are much smaller than the ones on the Siegel threefold. We explain how to compute even smaller polynomials by using pullbacks of Theta functions to the Hilbert surface, and give an application to the CRT method to construct class polynomials.

7.14. Individual Logarithm Step in Non-prime Fields

Participant: Aurore Guillevic.

In [20], the previous work [33] on speeding-up the first phase of the individual discrete logarithm computation, the initial splitting, a.k.a. smoothing phase, is extended to any non-prime finite field $\mathbb{F}_{p^n}$ where n is composite. It is also applied to the new variant Tower-NFS.

7.15. Last Year Results that Appeared in 2017

Our work [6], in collaboration with J. Fried and N. Heninger from the University of Pennsylvania, describing a kilobit discrete logarithm computation for a trapdoored prime number has been published in Eurocrypt 2017.

A paper detailing the implementation of the ECM factoring algorithm on the Kalray MPPA-256 many-core processor, written as a collaboration between Jérémie Detrey and Pierrick Gaudry from CARAMBA, and Masahiro Ishii, Atsuo Inomata, and Kazutoshi Fujikawa from NAIST (Nara, Japan), was published in IEEE Transaction on Computers [2].

In [39], the notions of Square, saturation, integrals, multisets, bit patterns and tuples cryptanalysis are revised. A new Slice & Fuse paradigm to better exploit multiset type properties of block ciphers is proposed. With this refined analysis, we improve the best bounds proposed in such contexts against the following block ciphers: Threefish, Prince, Present and Rectangle.

In [3], we improve the existing impossible-differential attacks against Rijndael-160 and Rijndael-224.

Our work [10] about the computational power of the Measurement-based Quantum Computation model, written by Luc Sanselme and Simon Perdrix (from the CARTE team at LORIA), has appeared.

CASCADE Project-Team

6. New Results

6.1. Results

All the results of the team have been published in journals or conferences (see the list of publications). They are all related to the research program (see before) and the research projects (see after):

- More efficient constructions with lattices
- New e-cash constructions
- Advanced primitives for the privacy in the cloud
- Efficient functional encryption
- Various predicate encryption schemes

DATASHAPE Project-Team

7. New Results

7.1. Algorithmic aspects of topological and geometric data analysis

7.1.1. Variance Minimizing Transport Plans for Inter-surface Mapping

Participant: David Cohen-Steiner.

In collaboration with Manish Mandad, Leik Kobbelt (RWTH Aachen), Pierre Alliez (Inria), and Mathieu Desbrun (Caltech).

We introduce an effcient computational method for generating dense and low distortion maps between two arbitrary surfaces of same genus. Instead of relying on semantic correspondences or surface parameterization, we directly optimize a variance-minimizing transport plan between two input surfaces that defines an as-conformal-as-possible inter-surface map satisfying a user-prescribed bound on area distortion. The transport plan is computed via two alternating convex optimizations, and is shown to minimize a generalized Dirichlet energy of both the map and its inverse. Computational efficiency is achieved through a coarse-tone approach in diffusion geometry, with Sinkhorn iterations modified to enforce bounded area distortion. The resulting inter-surface mapping algorithm applies to arbitrary shapes robustly, with little to no user interaction.

7.1.2. Approximating the spectrum of a graph

Participant: David Cohen-Steiner.

In collaboration with Weihao Kong, Gregory Valiant (Stanford), and Christian Sohler (TU Dortmund).

The spectrum of a network or graph $G = (V, E)$ with adjacency matrix A consists of the eigenvalues of the normalized Laplacian $L = I - D^{-1/2}AD^{-1/2}$. This set of eigenvalues encapsulates many aspects of the structure of the graph, including the extent to which the graph posses community structures at multiple scales. We study the problem of approximating the spectrum $\lambda = (\lambda_1, \ldots, \lambda_{|V|})$, $0 \leq \lambda 1, \leq \ldots, \leq \lambda |V| \leq 2$ of G in the regime where the graph is too large to explicitly calculate the spectrum. We present a sublinear time algorithm that, given the ability to query a random node in the graph and select a random neighbor of a given node, computes a succinct representation of an approximation $\widetilde{\lambda}$ such that $\|\widetilde{\lambda} - \lambda\|_1 \leq \varepsilon |V|$. Our algorithm has query complexity and running time $\exp(O(1/\varepsilon))$, independent of the size of the graph, $|V|$. We demonstrate the practical viability of our algorithm on 15 different real-world graphs from the Stanford Large Network Dataset Collection, including social networks, academic collaboration graphs, and road networks. For the smallest of these graphs, we are able to validate the accuracy of our algorithm by explicitly calculating the true spectrum; for the larger graphs, such a calculation is computationally prohibitive. In addition we study the implications of our algorithm to property testing in the bounded degree graph model.

7.1.3. Anisotropic triangulations via discrete Riemannian Voronoi diagrams

Participants: Jean-Daniel Boissonnat, Mathijs Wintraecken.

In collaboration with mael Rouxel-Labbé (GeometryFactory).

The construction of anisotropic triangulations is desirable for various applications, such as the numerical solving of partial differential equations and the representation of surfaces in graphics. To solve this notoriously difficult problem in a practical way, we introduce the discrete Riemannian Voronoi diagram, a discrete structure that approximates the Riemannian Voronoi diagram. This structure has been implemented and was shown to lead to good triangulations in $\mathbb{R}^2$ and on surfaces embedded in $\mathbb{R}^3$ as detailed in our experimental companion paper.

In [23], [32], [34], we study theoretical aspects of our structure. Given a finite set of points $\mathcal{P}$ in a domain Ω equipped with a Riemannian metric, we compare the discrete Riemannian Voronoi diagram of $\mathcal{P}$ to its Riemannian Voronoi diagram. Both diagrams have dual structures called the discrete Riemannian Delaunay and the Riemannian Delaunay complex. We provide conditions that guarantee that these dual structures are identical. It then follows from previous results that the discrete Riemannian Delaunay complex can be embedded in Ω under sufficient conditions, leading to an anisotropic triangulation with curved simplices. Furthermore, we show that, under similar conditions, the simplices of this triangulation can be straightened.

7.1.4. Only distances are required to reconstruct submanifolds

Participants: Jean-Daniel Boissonnat, Ramsay Dyer, Steve Oudot.

In collaboration with Arijit Ghosh (Indian Statistical Institute).

In [14], we give the first algorithm that outputs a faithful reconstruction of a submanifold of Euclidean space without maintaining or even constructing complicated data structures such as Voronoi diagrams or Delaunay complexes. Our algorithm uses the witness complex and relies on the stability of *power protection*, a notion introduced in this paper. The complexity of the algorithm depends exponentially on the intrinsic dimension of the manifold, rather than the dimension of ambient space, and linearly on the dimension of the ambient space. Another interesting feature of this work is that no explicit coordinates of the points in the point sample is needed. The algorithm only needs the *distance matrix* as input, i.e., only distance between points in the point sample as input.

7.1.5. An obstruction to Delaunay triangulations in Riemannian manifolds

Participants: Jean-Daniel Boissonnat, Ramsay Dyer.

In collaboration with Arijit Ghosh (Indian Statistical Institute) and Nikolay Martynchuk (University of Groningen).

Delaunay has shown that the Delaunay complex of a finite set of points P of Euclidean space $\mathbb{R}^m$ triangulates the convex hull of P, provided that P satisfies a mild genericity property. Voronoi diagrams and Delaunay complexes can be defined for arbitrary Riemannian manifolds. However, Delaunay's genericity assumption no longer guarantees that the Delaunay complex will yield a triangulation; stronger assumptions on P are required. A natural one is to assume that P is sufficiently dense. Although results in this direction have been claimed, we show that sample density alone is insufficient to ensure that the Delaunay complex triangulates a manifold of dimension greater than 2 [13].

7.1.6. Local criteria for triangulation of manifolds

Participants: Jean-Daniel Boissonnat, Ramsay Dyer, Mathijs Wintraecken.

In collaboration with Arijit Ghosh (Indian Statistical Institute).

We present criteria for establishing a triangulation of a manifold [40]. Given a manifold M, a simplicial complex $\mathcal{A}$, and a map H from the underlying space of $\mathcal{A}$ to M, our criteria are presented in local coordinate charts for M, and ensure that H is a homeomorphism. These criteria do not require a differentiable structure, or even an explicit metric on M. No Delaunay property of $\mathcal{A}$ is assumed. The result provides a triangulation guarantee for algorithms that construct a simplicial complex by working in local coordinate patches. Because the criteria are easily checked algorithmically, they are expected to be of general use.

7.1.7. Triangulating stratified manifolds I: a reach comparison theorem

Participants: Jean-Daniel Boissonnat, Mathijs Wintraecken.

In [42], we define the reach for submanifolds of Riemannian manifolds, in a way that is similar to the Euclidean case. Given a d-dimensional submanifold $\mathcal{S}$ of a smooth Riemannian manifold $\mathbb{M}$ and a point $p \in \mathbb{M}$ that is not too far from $\mathcal{S}$ we want to give bounds on local feature size of $\exp_p^{-1}(\mathcal{S})$. Here $\exp_p^{-1}$ is the inverse exponential map, a canonical map from the manifold to the tangent space. Bounds on the local feature size of $\exp_p^{-1}(\mathcal{S})$ can be reduced to giving bounds on the reach of $\exp_p^{-1}(\mathcal{B})$, where $\mathcal{B}$ is a geodesic ball, centred at c with radius equal to the reach of $\mathcal{S}$. Equivalently we can give bounds on the reach of $\exp_p^{-1} \circ \exp_c(B_c)$, where now B_c is a ball in the tangent space $T_c\mathbb{M}$, with the same radius. To establish bounds on the reach of $\exp_p^{-1} \circ \exp_c(B_c)$ we use bounds on the metric and on its derivative in Riemann normal coordinates.

This result is a first step towards answering the important question of how to triangulate stratified manifolds.

7.1.8. The reach, metric distortion, geodesic convexity and the variation of tangent spaces

Participants: Jean-Daniel Boissonnat, Mathijs Wintraecken.

In collaboration with André Lieutier (Dassault Système).

In [41], we discuss three results. The first two concern general sets of positive reach: We first characterize the reach by means of a bound on the metric distortion between the distance in the ambient Euclidean space and the set of positive reach. Secondly, we prove that the intersection of a ball with radius less than the reach with the set is geodesically convex, meaning that the shortest path between any two points in the intersection lies itself in the intersection. For our third result we focus on manifolds with positive reach and give a bound on the angle between tangent spaces at two different points in terms of the distance between the points and the reach.

7.1.9. Delaunay triangulation of a random sample of a good sample has linear size

Participants: Jean-Daniel Boissonnat, Kunal Dutta, Marc Glisse.

In collaboration with Olivier Devillers (Inria Nancy Grand Est).

The *randomized incremental construction* (RIC) for building geometric data structures has been analyzed extensively, from the point of view of worst-case distributions. In many practical situations however, we have to face nicer distributions. A natural question that arises is: do the usual RIC algorithms automatically adapt when the point samples are nicely distributed. We answer positively to this question for the case of the Delaunay triangulation of ϵ-nets.

ϵ-nets are a class of nice distributions in which the point set is such that any ball of radius ϵ contains at least one point of the net and two points of the net are distance at least ϵ apart. The Delaunay triangulations of ϵ-nets are proved to have linear size; unfortunately this is not enough to ensure a good time complexity of the randomized incremental construction of the Delaunay triangulation. In [33], [38], we prove that a uniform random sample of a given size that is taken from an ϵ-net has a linear sized Delaunay triangulation in any dimension. This result allows us to prove that the randomized incremental construction needs an expected linear size and an expected $O(n \log n)$ time.

Further, we also prove similar results in the case of non-Euclidean metrics, when the point distribution satisfies a certain *bounded expansion* property; such metrics can occur, for example, when the points are distributed on a low-dimensional manifold in a high-dimensional ambient space.

7.1.10. Kernelization of the Subset General Position problem in Geometry

Participants: Jean-Daniel Boissonnat, Kunal Dutta.

In collaboration with Arijit Ghosh (Indian Statistical Institute) and Sudeshna Kolay (Eindhoven University of Technology).

In [21], we consider variants of the GEOMETRIC SUBSET GENERAL POSITION problem. In defining this problem, a geometric subsystem is specified, like a subsystem of lines, hyperplanes or spheres. The input of the problem is a set of n points in $\mathbb{R}^d$ and a positive integer k. The objective is to find a subset of at least k input points such that this subset is in general position with respect to the specified subsystem. For example, a set of points is in general position with respect to a subsystem of hyperplanes in $\mathbb{R}^d$ if no $d+1$ points lie on the same hyperplane. In this paper, we study the HYPERPLANE SUBSET GENERAL POSITION problem under two parameterizations. When parameterized by k then we exhibit a polynomial kernelization for the problem. When parameterized by $h = n - k$, or the dual parameter, then we exhibit polynomial kernels which are also tight, under standard complexity theoretic assumptions. We can also conclude similar kernelization results for D-POLYNOMIAL SUBSET GENERAL POSITION, where a vector space of polynomials of degree at most d are specified as the underlying subsystem such that the size of the basis for this vector space is b. The objective is to find a set of at least k input points, or in the dual delete at most $h = n - k$ points, such that no $b+1$ points lie on the same polynomial. Notice that this is a generalization of many well-studied geometric variants of the SET COVER problem, such as CIRCLE SUBSET GENERAL POSITION. We also study the general projective variants of these problems. These problems are also related to other geometric problems like SUBSET DELAUNAY TRIANGULATION problem.

7.1.11. Tight Kernels for Covering and Hitting: Point Hyperplane Cover and Polynomial Point Hitting Set

Participants: Jean-Daniel Boissonnat, Kunal Dutta.

In collaboration with Arijit Ghosh (Indian Statistical Institute) and Sudeshna Kolay (Eindhoven University of Technology).

The POINT HYPERPLANE COVER problem in $\mathbb{R}^d$ takes as input a set of n points in $\mathbb{R}^d$ and a positive integer k. The objective is to cover all the given points with a set of at most k hyperplanes. The D-POLYNOMIAL POINTS HITTING SET (D-POLYNOMIAL POINTS HS) problem in $\mathbb{R}^d$ takes as input a family $\mathcal{F}$ of D-degree polynomials from a vector space $\mathcal{R}$ in $\mathbb{R}^d$, and determines whether there is a set of at most k points in $\mathbb{R}^d$ that hit all the polynomials in $\mathcal{F}$. In [22], we exhibit tight kernels where k is the parameter for these problems.

7.1.12. Shallow packings, semialgebraic set systems, Macbeath regions, and polynomial partitioning

Participant: Kunal Dutta.

In collaboration with Arijit Ghosh (Indian Statistical Institute) and Bruno Jartoux (Université Paris-Est, Laboratoire d'Informatique Gaspard-Monge, ESIEE Paris, France) and Nabil H. Mustafa (Université Paris-Est, Laboratoire d'Informatique Gaspard-Monge, ESIEE Paris, France).

The packing lemma of Haussler states that given a set system $(X, \mathbb{R})$ with bounded VC dimension, if every pair of sets in $\mathbb{R}$ have large symmetric difference, then $\mathbb{R}$ cannot contain too many sets. Recently it was generalized to the shallow packing lemma, applying to set systems as a function of their shallow-cell complexity. In [29] we present several new results and applications related to packings:

1. an optimal lower bound for shallow packings,
2. improved bounds on Mnets, providing a combinatorial analogue to Macbeath regions in convex geometry,
3. we observe that Mnets provide a general, more powerful framework from which the state-of-the-art unweighted ϵ-net results follow immediately, and
4. simplifying and generalizing one of the main technical tools in Fox *et al.* (*J. of the EMS*, to appear).

7.1.13. A Simple Proof of Optimal Epsilon Nets

Participant: Kunal Dutta.

In collaboration with Nabil H. Mustafa (Université Paris-Est, Laboratoire d'Informatique Gaspard-Monge, ESIEE Paris, France, and Arijit Ghosh (Indian Statistical Institute)).

Showing the existence of ϵ-nets of small size has been the subject of investigation for almost 30 years, starting from the initial breakthrough of Haussler and Welzl (1987). Following a long line of successive improvements, recent results have settled the question of the size of the smallest ϵ-nets for set systems as a function of their so-called shallow-cell complexity.

In [20] we give a short proof of this theorem in the space of a few elementary paragraphs, showing that it follows by combining the ϵ-net bound of Haussler and Welzl (1987) with a variant of Haussler's packing lemma (1991).

This implies all known cases of results on unweighted ϵ-nets studied for the past 30 years, starting from the result of Matoušek, Seidel and Welzl (1990) to that of Clarkson and Varadajan (2007) to that of Varadarajan (2010) and Chan, Grant, Könemann and Sharpe (2012) for the unweighted case, as well as the technical and intricate paper of Aronov, Ezra and Sharir (2010).

7.1.14. On Subgraphs of Bounded Degeneracy in Hypergraphs

Participant: Kunal Dutta.

In collaboration with Arijit Ghosh (Indian Statistical Institute)).

A k-uniform hypergraph is d-degenerate if every induced subgraph has a vertex of degree at most d. In [48], given a k-uniform hypergraph $H = (V(H), E(H))$, we show there exists an induced subgraph of size at least

$$\sum_{v \in V(H)} \min \left\{ 1,\, c_k \left(\frac{d+1}{d_H(v)+1} \right)^{1/(k-1)} \right\},$$

where $c_k = 2^{-\left(1+\frac{1}{k-1}\right)} \left(1 - \frac{1}{k}\right)$ and $d_H(v)$ denotes the degree of vertex v in the hypergraph H. This connects, extends, and generalizes results of Alon-Kahn-Seymour (1987), on d-degenerate sets of graphs, Dutta-Mubayi-Subramanian (2012) on d-degenerate sets of linear hypergraphs, and Srinivasan-Shachnai (2004) on independent sets in hypergraphs to d-degenerate subgraphs of hypergraphs. Our technique also gives optimal lower bounds for a more generalized definition of degeneracy introduced by Zaker (2013). We further give a simple non-probabilistic proof of the Dutta-Mubayi-Subramanian bound for linear k-uniform hypergraphs, which extends the Alon-Kahn-Seymour proof technique to hypergraphs. Finally we provide several applications in discrete geometry, extending results of Payne-Wood (2013) and Cardinal-Tóth-Wood (2016). We also address some natural algorithmic questions. The proof of our main theorem combines the *random permutation* technique of Bopanna-Caro-Wei and Beame and Luby, together with a new *local density* argument which may be of independent interest.

7.2. Statistical aspects of topological and geometric data analysis

7.2.1. The DTM-signature for a geometric comparison of metric-measure spaces from samples

Participant: Claire Brécheteau.

In [43], we introduce the notion of DTM-signature, a measure on $\mathbb{R}_+$ that can be associated to any metric-measure space. This signature is based on the distance to a measure (DTM) introduced by Chazal, Cohen-Steiner and Mérigot. It leads to a pseudo-metric between metric-measure spaces, upper-bounded by the Gromov-Wasserstein distance. Under some geometric assumptions, we derive lower bounds for this pseudo-metric. Given two N-samples, we also build an asymptotic statistical test based on the DTM-signature, to reject the hypothesis of equality of the two underlying metric measure spaces, up to a measure-preserving isometry. We give strong theoretical justifications for this test and propose an algorithm for its implementation.

7.2.2. Estimating the Reach of a Manifold

Participants: Eddie Aamari, Frédéric Chazal, Bertrand Michel.

In collaboration with J. Kim, A. Rinaldo, L. Wasserman (Carnegie Mellon University)

Various problems of computational geometry and manifold learning encode geometric regularity through the so-called reach, a generalized convexity parameter. The reach τ_M of a submanifold $M \subset \mathbb{R}^D$ is the maximal offset radius on which the projection onto M is well defined. The quantity τ_M renders a certain minimal scale of M, giving bounds on both maximum curvature and possible bottleneck structures. In [35], we study the geometry of the reach through an approximation perspective. We derive new geometric results on the reach for submanifolds without boundary. An estimator $\widehat{\tau}$ of τ_M is proposed in a framework where tangent spaces are known, and bounds assessing its efficiency are derived. In the case of i.i.d. random point cloud $\mathbb{X}_n$, $\widehat{\tau}(\mathbb{X}_n)$ is showed to achieve uniform expected loss bounds over a $\mathcal{C}^3$-like model. Minimax upper and lower bounds are derived, and we conclude with the extension to a model with unknown tangent spaces.

7.2.3. Robust Topological Inference: Distance To a Measure and Kernel Distance

Participants: Frédéric Chazal, Bertrand Michel.

In collaboration with B. Fasy, F. Lecci, A. Rinaldo, L. Wasserman.

Let P be a distribution with support S. The salient features of S can be quantified with persistent homology, which summarizes topological features of the sublevel sets of the distance function (the distance of any point x to S). Given a sample from P we can infer the persistent homology using an empirical version of the distance function. However, the empirical distance function is highly non-robust to noise and outliers. Even one outlier is deadly. The distance-to-a-measure (DTM) and the kernel distance are smooth functions that provide useful topological information but are robust to noise and outliers. In [17], we derive limiting distributions and confidence sets, and we propose a method for choosing tuning parameters.

7.2.4. Statistical analysis and parameter selection for Mapper

Participants: Steve Oudot, Bertrand Michel, Mathieu Carrière.

In [44] we study the question of the statistical convergence of the 1-dimensional Mapper to its continuous analogue, the Reeb graph. We show that the Mapper is an optimal estimator of the Reeb graph, which gives, as a byproduct, a method to automatically tune its parameters and compute confidence regions on its topological features, such as its loops and flares. This allows to circumvent the issue of testing a large grid of parameters and keeping the most stable ones in the brute-force setting, which is widely used in visualization, clustering and feature selection with the Mapper.

7.2.5. Sliced Wasserstein Kernel for Persistence Diagrams

Participants: Steve Oudot, Mathieu Carrière.

In collaboration with M. Cuturi (ENSAE)

Persistence diagrams (PDs) play a key role in topological data analysis (TDA), in which they are routinely used to describe succinctly complex topological properties of complicated shapes. PDs enjoy strong stability properties and have proven their utility in various learning contexts. They do not, however, live in a space naturally endowed with a Hilbert structure and are usually compared with specific distances, such as the bottleneck distance. To incorporate PDs in a learning pipeline, several kernels have been proposed for PDs with a strong emphasis on the stability of the RKHS distance w.r.t. perturbations of the PDs. In [27], we use the Sliced Wasserstein approximation of the Wasserstein distance to define a new kernel for PDs, which is not only provably stable but also provably discriminative w.r.t. the Wasserstein distance W1∞ between PDs. We also demonstrate its practicality, by developing an approximation technique to reduce kernel computation time, and show that our proposal compares favorably to existing kernels for PDs on several benchmarks.

7.2.6. An introduction to Topological Data Analysis: fundamental and practical aspects for data scientists

Participants: Frédéric Chazal, Bertrand Michel.

Topological Data Analysis (TDA) is a recent and fast growing field providing a set of new topological and geometric tools to infer relevant features for possibly complex data. In [45], we propose a brief introduction, through a few selected recent and state-of-the-art topics, to basic fundamental and practical aspects of TDA for non experts.

7.3. Topological approach for multimodal data processing

7.3.1. On the Stability of Functional Maps and Shape Difference Operators

Participants: Frédéric Chazal, Ruqi Huang, Maks Ovsjanikov.

In this paper, we provide stability guarantees for two frameworks that are based on the notion of functional maps. We consider two types of perturbations in our analysis: one is on the input shapes and the other is on the change in *scale*. In theory, we formulate and justify the robustness that has been observed in practical implementations of those frameworks. Inspired by our theoretical results, we propose a pipeline for constructing shape difference operators on point clouds and show numerically that the results are robust and informative. In particular, we show that both the shape difference operators and the derived areas of highest distortion are stable with respect to changes in shape representation and change of scale. Remarkably, this is in contrast with the well-known instability of the eigenfunctions of the Laplace-Beltrami operator computed on point clouds compared to those obtained on triangle meshes.

7.3.2. Local Equivalence and Intrinsic Metrics Between Reeb Graphs

Participants: Steve Oudot, Mathieu Carrière.

As graphical summaries for topological spaces and maps, Reeb graphs are common objects in the computer graphics or topological data analysis literature. Defining good metrics between these objects has become an important question for applications, where it matters to quantify the extent by which two given Reeb graphs differ. Recent contributions emphasize this aspect, proposing novel distances such as functional distortion or interleaving that are provably more discriminative than the so-called bottleneck distance, being true metrics whereas the latter is only a pseudo-metric. Their main drawback compared to the bottleneck distance is to be comparatively hard (if at all possible) to evaluate. In [28] we take the opposite view on the problem and show that the bottleneck distance is in fact good enough locally, in the sense that it is able to discriminate a Reeb graph from any other Reeb graph in a small enough neighborhood, as efficiently as the other metrics do. This suggests considering the intrinsic metrics induced by these distances, which turn out to be all globally equivalent. This novel viewpoint on the study of Reeb graphs has a potential impact on applications, where one may not only be interested in discriminating between data but also in interpolating between them.

7.3.3. Structure and Stability of the One-Dimensional Mapper

Participants: Steve Oudot, Mathieu Carrière.

Given a continuous function $f : X -> R$ and a cover I of its image by intervals, the Mapper is the nerve of a refinement of the pullback cover $f^{-1}(I)$. Despite its success in applications, little is known about the structure and stability of this construction from a theoretical point of view. As a pixelized version of the Reeb graph of f, it is expected to capture a subset of its features (branches, holes), depending on how the interval cover is positioned with respect to the critical values of the function. Its stability should also depend on this positioning. In [16] we propose a theoretical framework relating the structure of the Mapper to that of the Reeb graph, making it possible to predict which features will be present and which will be absent in the Mapper given the function and the cover, and for each feature, to quantify its degree of (in-)stability. Using this framework, we can derive guarantees on the structure of the Mapper, on its stability, and on its convergence to the Reeb graph as the granularity of the cover I goes to zero.

7.4. Experimental research and software development

7.4.1. Stride detection for pedestrian trajectory reconstruction: a machine learning approach based on geometric patterns

Participants: Frédéric Chazal, Bertrand Michel, Bertrand Beaufils.

In collaboration with M. Grelet (Sysnav)

A strides detection algorithm is proposed using inertial sensors worn on the ankle. This innovative approach based on geometric patterns can detect both normal walking strides and atypical strides such as small steps, side steps and backward walking that existing methods struggle to detect. It is also robust in critical situations, when for example the wearer is sitting and moving the ankle, while most algorithms in the literature would wrongly detect strides.

GAMBLE Project-Team

7. New Results

7.1. Non-Linear Computational Geometry

Participants: Sény Diatta, Laurent Dupont, George Krait, Sylvain Lazard, Guillaume Moroz, Marc Pouget.

7.1.1. Reliable location with respect to the projection of a smooth space curve

Consider a plane curve $\mathcal{B}$ defined as the projection of the intersection of two analytic surfaces in $\mathbb{R}^3$ or as the apparent contour of a surface. In general, $\mathcal{B}$ has node or cusp singular points and thus is a singular curve. Our main contribution [9] is the computation of a data structure for answering point location queries with respect to the subdivision of the plane induced by $\mathcal{B}$. This data structure is composed of an approximation of the space curve together with a topological representation of its projection $\mathcal{B}$. Since $\mathcal{B}$ is a singular curve, it is challenging to design a method only based on reliable numerical algorithms.

In a previous work [39], we have shown how to describe the set of singularities of $\mathcal{B}$ as regular solutions of a so-called ball system suitable for a numerical subdivision solver. Here, the space curve is first enclosed in a set of boxes with a certified path-tracker to restrict the domain where the ball system is solved. Boxes around singular points are then computed such that the correct topology of the curve inside these boxes can be deduced from the intersections of the curve with their boundaries. The tracking of the space curve is then used to connect the smooth branches to the singular points. The subdivision of the plane induced by $\mathcal{B}$ is encoded as an extended planar combinatorial map allowing point location. We experimented our method and show that our reliable numerical approach can handle classes of examples that are not reachable by symbolic methods.

7.1.2. Computing effectively stabilizing controllers for a class of nD systems

In this paper [1], we study the internal stabilizability and internal stabilization problems for multidimensional (nD) systems. Within the fractional representation approach, a multidimensional system can be studied by means of matrices with entries in the integral domain of structurally stable rational fractions, namely the ring of rational functions which have no poles in the closed unit polydisc $\overline{\mathbb{U}}^n = \{z = (z_1, ..., z_n) \in \mathbb{C}^n \mid |z_1| \leq 1, ..., |z_n| \leq 1\}$.

It is known that the internal stabilizability of a multidimensional system can be investigated by studying a certain polynomial ideal $I = \langle p_1, ..., p_r \rangle$ that can be explicitly described in terms of the transfer matrix of the plant. More precisely the system is stabilizable if and only if $V(I) = \{z \in \mathbb{C}^n \mid p_1(z) = \cdots = p_r(z) = 0\} \cap \overline{\mathbb{U}}^n = \varnothing$. In the present article, we consider the specific class of linear nD systems (which includes the class of 2D systems) for which the ideal I is zero-dimensional, i.e., the p_i's have only a finite number of common complex zeros. We propose effective symbolic-numeric algorithms for testing if $V(I) \cap \overline{\mathbb{U}}^n = \varnothing$, as well as for computing, if it exists, a stable polynomial $p \in I$ which allows the effective computation of a stabilizing controller. We illustrate our algorithms through an example and finally provide running times of prototype implementations for 2D and 3D systems.

7.2. Non-Euclidean Computational Geometry

Participants: Vincent Despré, Iordan Iordanov, Monique Teillaud.

7.2.1. Implementing Delaunay Triangulations of the Bolza Surface

The CGAL library offers software packages to compute Delaunay triangulations of the (flat) torus of genus one in two and three dimensions. To the best of our knowledge, there is no available software for the simplest possible extension, i.e., the Bolza surface, a hyperbolic manifold homeomorphic to a torus of genus two. We present an implementation based on the theoretical results and the incremental algorithm proposed recently. We describe the representation of the triangulation, we detail the different steps of the algorithm, we study predicates, and report experimental results [5]. The implementation is publicly available in the development branch of CGAL on `github`[0] and will soon be submitted for integration in the library.

[0] https://members.loria.fr/Monique.Teillaud/DT_Bolza_SoCG17/

7.3. Probabilistic Analysis of Geometric Data Structures and Algorithms

Participants: Olivier Devillers, Charles Duménil.

7.3.1. Delaunay triangulation of a random sample of a good sample has linear size

A good sample is a point set such that any ball of radius ϵ contains a constant number of points. The Delaunay triangulation of a good sample is proved to have linear size, unfortunately this is not enough to ensure a good time complexity of the randomized incremental construction of the Delaunay triangulation. In this paper we prove that a random Bernoulli sample of a good sample has a triangulation of linear size. This result allows to prove that the randomized incremental construction needs an expected linear size and an expected $O(n \log n)$ time [8].

This work was done in collaboration with Marc Glisse (Project-team DATASHAPE).

7.3.2. Delaunay triangulation of a random sampling of a generic surface

The complexity of the Delaunay triangulation of n points distributed on a surface ranges from linear to quadratic. We prove that when the points are evenly distributed on a smooth compact generic surface the expected size of the Delaunay triangulation is $O(n)$. This result has to be compared with a bound of $O(n \log n)$ when the points are a deterministic good sample of the surface under the same hypotheses on the surface [13].

7.4. Classical Computational Geometry and Graph Drawing

Participants: Olivier Devillers, Sylvain Lazard.

7.4.1. Celestial Walk: A Terminating Oblivious Walk for Convex Subdivisions

We present a new oblivious walking strategy for convex subdivisions. Our walk is faster than the straight walk and more generally applicable than the visiblity walk. To prove termination of our walk we use a novel monotonically decreasing distance measure [10].

This work was done in collaboration with Wouter Kuijper and Victor Ermolaev (Nedap Security Management).

7.4.2. Snap rounding polyhedral subdivisions

Let $\mathcal{P}$ be a set of n polygons in $\mathbb{R}^3$, each of constant complexity and with pairwise disjoint interiors. We propose a rounding algorithm that maps $\mathcal{P}$ to a simplicial complex $\mathcal{Q}$ whose vertices have integer coordinates. Every face of $\mathcal{P}$ is mapped to a set of faces (or edges or vertices) of $\mathcal{Q}$ and the mapping from $\mathcal{P}$ to $\mathcal{Q}$ can be build through a continuous motion of the faces such that (i) the L_∞ Hausdorff distance between a face and its image during the motion is at most 3/2 and (ii) if two points become equal during the motion they remain equal through the rest of the motion. In the worse, the size of $\mathcal{Q}$ is $O(n^{15})$, but, under reasonable hypotheses, this complexities decreases to $O(n^5)$.

This work was done in collaboration with William J. Lenhart (Williams College, USA).

7.4.3. Explicit array-based compact data structures for triangulations

We consider the problem of designing space efficient solutions for representing triangle meshes. Our main result is a new explicit data structure for compactly representing planar triangulations: if one is allowed to permute input vertices, then a triangulation with n vertices requires at most $4n$ references ($5n$ references if vertex permutations are not allowed). Our solution combines existing techniques from mesh encoding with a novel use of maximal Schnyder woods. Our approach extends to higher genus triangulations and could be applied to other families of meshes (such as quadrangular or polygonal meshes). As far as we know, our solution provides the most parsimonious data structures for triangulations, allowing constant time navigation. Our data structures require linear construction time, and are fast decodable from a standard compressed format without using additional memory allocation. All bounds, concerning storage requirements and navigation performances, hold in the worst case. We have implemented and tested our results, and experiments confirm the practical interest of compact data structures.

This work was done in collaboration with Luca Castelli Aleardi (LIX).

GRACE Project-Team

6. New Results

6.1. qDSA: Compact signatures for IoT

B. Smith and Joost Renes (Radboud University, NL) developed **qDSA**, a new digital signature scheme targeting constrained devices, typically microcontrollers with extremely limited memory. An article describing qDSA was presented at ASIACRYPT 2017, and a reference implementation software package has been placed into the public domain.

6.2. PIR based on transversal designs

J. Lavauzelle presented a construction of Private Information Retrieval (PIR) protocols from combinatorial structures called transversal designs. The construction features low computation and low storage overhead for the servers. For some instances, adequate communication between servers and user is achieved. The PIR scheme also generalizes to colluding servers. The construction has been presented during WCC 2017 [17], and in a poster session in the Munich Workshop in Coding and Applications.

6.3. On the security of compact McEliece keys

E. Barelli presented at WCC 2017 (Workshop on Coding and Cryptography, St Petersburg, Russia) her recent results on the analysis of McEliece scheme based on alternant codes with a non trivial automorphism group [16]. These codes were suggested for public key encryption since, compared to codes with trivial automorphism group, they could provide shorter keys.

If the security with respect to generic decoding attacks is almost unchanged when considering codes with non trivial automorphisms, E. Barelli proved that the security with respect to key recovery attacks is highly reduced since, it reduces to recover the structure of the subcode of fixed elements by the automorphism group.

6.4. Two-points codes on the generalized Giuletti Korchmaros curve

In a collaboration with Peter Beelen, Mrinmoy Datta, Vincent Neiger and Johan Rosenkilde (DTU Copenhagen), E. Barelli obtained improved lower bounds for the minimum distance of some algebraic geometry codes from Giulietti Korchmaros curves [20].

6.5. Towards a function field version of Freiman's theorem

In a collaboration with Christine Bachoc and Gilles Zémor (University of Bordeaux), A. Couvreur obtained a characterisation of subspaces S of a function field F over an algebraically closed field satisfiying

$$\dim S^2 = 2 \dim S$$

where S^2 denotes the space spanned by all the products of two elements of S. They obtained the following result [18]:

Theorem. *Let F be a function field over an algebraically closed field, and S be a finite dimensional subspace of F which spans F as an algebra and such that*

$$\dim S^2 = 2 \dim S.$$

Then F is a function field of transcendence degree 1 and

- *either F has genus 1 and S is a Riemann Roch space*
- *or F has genus 0 and S is a subspace of codimension 1 in a Riemann Roch space.*

6.6. BIG QUAKE

In the context of NIST's call for post quantum cryptosystems:

https://csrc.nist.gov/Projects/Post-Quantum-Cryptography

A. Couvreur and E. Barelli participated to the submission BIG QUAKE [19] (BInary Goppa QUAsi–cyclic Key Encapsulation). The proposal consists in a public key encryption scheme (with a conversion to a Key Encapsulation Mechanism) using binary quasi–cyclic Goppa codes.

The details on the proposal are on the following website.

https://bigquake.inria.fr/

6.7. Discrete Logarithm computations in finite fields with the NFS algorithm

The best discrete logarithm record computations in prime fields and large characteristic finite fields are obtained with Number Field Sieve algorithm (NFS) at the moment.

6.7.1. Computing discrete logarithms in $GF(p^6)$

A. Guillevic, L. Grémy, F. Morain and E. Thomé (from CARAMBA EPC in LORIA) computed a discrete log on a curve of embedding degree 6 and cryptographic size. This clearly showed that curves with small embedding degrees are indeed weak. The article [23] was presented by L. Grémy during the SAC 2017 conference in Ottawa.

6.7.2. Identity management on Bitcoin's blockchain

D. Augot and W. George in collaboration with Hervé Chabanne (Safran Identity and Security, ex Morpho, now Idemia) designed two schemes to allow users to authenticate using so-called anonymous credentials, issues by an identity provider. We used Brands anonymous credentials with selective disclosure each time, first for a finely tuned, user managed, identity scheme [12], second for a more classical high throughput scheme [13], inspired by CONIKS https://coniks.cs.princeton.edu.

6.7.3. Law and Blokchain smart contracts

D. Augot, with Célia Zolynski, is co-advising Hanna-Mae Bisserier, a PhD student law, on the impact of blockchains on legal systems. The PhD is in law, and D. Augot only gives scientific and technological explanations, while the direction of the thesis is done by Célia Zolynski.

LFANT Project-Team

6. New Results

6.1. Non commutative number theory

Participant: Jean Paul Cerri.

Pierre Lezowski has studied in [11], Euclidean properties of matrix algebras. He proved that if A is a commutative ring and if $n > 1$ is an integer , then $M_n(A)$ is right and left Euclidean if and only if A is a principal ideal ring. Moreover, under the hypothesis that the stathm takes integer values, he established that if A is an integral domain, then $M_n(A)$ is ω-stage right and left Euclidean if and only if A is a Bézout ring. He also proved, under the same hypothesis, that if A is a K-Hermite ring, then $M_n(A)$ is $(4n-3)$-stage left and right Euclidean, that if A is an elementary divisor ring, then $M_n(A)$ is $(2n-1)$-stage left and right Euclidean, and that if A is a principal ideal ring, then $M_n(A)$ is 2-stage right and left Euclidean. In each case, he obtained an explicit algorithm allowing to compute, among other things, right or left gcd in $M_n(A)$.

Jean-Paul Cerri and Pierre Lezowski have generalized in [18], Cerri's algorithm (for the computation of the upper part of the norm-Euclidean spectrum of a number field) to totally definite quaternion fields. This allowed them to establish the exact value of the norm-Euclidean minimum of many orders in totally definite quaternion fields over a quadratic number field. Before this work, nobody knew how to compute the exact value of such a minimum when the base number field has degree > 1. They also proved that the Euclidean minimum and the inhomogeneous minimum of orders in such quaternion fields are always equal and that moreover they are rational under the hypothesis that the base number field is not quadratic, which remains the only open case, as for real number fields.

In [12] Lezowski determines which cyclic field of degree d are norm-Euclidean for $d = 5, 7, 19, 31, 37, 43, 47, 59, 67, 71, 73, 79, 97$.

6.2. Cryptographic Protocols

Participant: Guilhem Castagnos.

In [15] G. Castagnos, L. Imbert, and F. Laguillaumie revisit a recent cryptographic primitive called *encryption switching protocols* (ESP). This primitive was introduced by Couteau, Peters and Pointcheval last year. It allows to switch ciphertexts between two encryption schemes. If such an ESP is built with two schemes that are respectively additively and multiplicatively homomorphic, it naturally gives rise to a secure 2-party computation protocol. It is thus perfectly suited for evaluating functions, such as multivariate polynomials, given as arithmetic circuits. Couteau et al. built an ESP to switch between Elgamal and Paillier encryptions which do not naturally fit well together. Consequently, they had to design a clever variant of Elgamal over $\mathbf{Z}/n\mathbf{Z}$ with a costly shared decryption.

In this work, Castagnos *et. al.* first present a conceptually simple generic construction for encryption switching protocols. Then, they give an efficient instantiation of our generic approach that uses two well-suited protocols, namely a variant of Elgamal in $\mathbf{Z}/p\mathbf{Z}$ and the Castagnos-Laguillaumie encryption which is additively homomorphic over $\mathbf{Z}/p\mathbf{Z}$. Among other advantages, this allows to perform all computations modulo a prime p instead of an RSA modulus. Overall, this solution leads to significant reductions in the number of rounds as well as the number of bits exchanged by the parties during the interactive protocols. They also show how to extend its security to the malicious setting.

This paper was presented at the CRYPTO Conference 2017, and is part of the ALAMBIC project.

6.3. Algorithmic number theory

Participant: Henri Cohen.

The book [17] by Henri Cohen on *Modular Forms: A Classical Approach* has been published. The theory of modular forms is a fundamental tool used in many areas of mathematics and physics. It is also a very concrete subject in itself and abounds with an amazing number of surprising identities. This comprehensive textbook, gives a complete picture of the classical aspects of the subject, with an emphasis on explicit formulas. Content include: elliptic functions and theta functions, the modular group, its subgroups, and general aspects of holomorphic and nonholomorphic modular forms, with an emphasis on explicit examples. The heart of the book is the classical theory developed by Hecke and continued up to the Atkin–Lehner–Li theory of newforms and including the theory of Eisenstein series, Rankin–Selberg theory, and a more general theory of theta series including the Weil representation. The final chapter also explores in some detail more general types of modular forms such as half-integral weight, Hilbert, Jacobi, Maass, and Siegel modular forms.

The article by Bill Allombert, Jean-Paul Allouche and Michel Mendès France on *Euler's divergent series and an elementary model in Statistical Physics* has been published in Statistical Physics Ars Mathematica Contemporanea. This article study the multiple integral of a multivariate exponential taken with respect either to the Lebesgue measure or to the discrete uniform Bernoulli measure. In the first case the integral is linked to Euler's everywhere divergent power series and its generalizations, while in the second case the integral is linked to a one-dimensional model of spin systems as encountered in physics.

Bill Allombert has worked with Nicolas Brisebarre and Alain Lasjaunias on *a two-valued sequence and related continued fractions in power series fields*. They explicitly describe a noteworthy transcendental continued fraction in the field of power series over $\mathbb{Q}$, having irrationality measure equal to 3. This continued fraction is a generating function of a particular sequence in the set $\{1, 2\}$.

In the Pari software, K. Belabas and H. Cohen have added an extensive new package `mf` for modular forms. This package allows to build spaces of classical modular form $M_k(\Gamma_0(N), \chi)$ where $2k \in \mathbb{Z}$ and perform standard tasks like finding bases, splitting the space using Hecke operators and the computation of eigenforms. It also solves important difficult problems: the computation of forms of weight 1, the realization of Shimura lifts as an explicit isomorphism between Kohnen's $+$-space $S_k^+(\Gamma 0(4N), \chi)$ and $S_{2k-1}(\Gamma_0(N), \chi^2)$ and the Fourier expansion of $f\ |_k\ \gamma$ for arbitrary f and arbitrary $\gamma \in \mathrm{GL}_2(\mathbb{Q})^+$, which includes as a special case the expansion of f at all cusps (where other modular form packages usualy deal with the expansion at infinity and the cusps reachable via Atkin-Lehner operators, e.g. all cusps in squarefree levels). The latter is especially important as it allows an explicit description of Atkin-Lehner operators, the evaluation of f arbitrary points in the upper-half plane, the computation of period polynomials and Pettersson products, etc.

6.4. Elliptic curve and Abelian varieties cryptology

Participant: Damien Robert.

In [21], E. Milio and D. Robert describe an algorithm to evaluate in quasi-linear time Hilbert modular functions in dimension 2, and also how to recover in time quasi-linear the period matrix from the value of the function. They apply this theory to the modular functions $j(\tau/\beta)$ and $\theta(\tau/\beta)$ where β is a totally real positive number of the quadratic real field corresponding to the Hilbert surface to construct modular polynomials parametrizing cyclic isogenies between principally polarised abelian varieties. This extends the construction of classical modular polynomials but allow to have much smaller polynomials, which allow to compute them up to norm $\ell = 91$ rather than $\ell = 7$ in dimension 2 for classical polynomials.

In [19], Dudeanu, Alina and Jetchev, Dimitar and Robert, Damien and Vuille, Marius describe an algorithm to compute cyclic isogenies from their kernels. This extends the work of [10] from isogenies with maximal isotropic kernels for the Weil pairing to cyclic isogenies, using real multiplication. Such isogenies are indispensable to fully explore the isogeny graph and will be able to speed up a lot of algorithms that needs isogenous curves, like the CRT method for class polynomials.

6.5. Arbitrary-precision ball arithmetic

Participant: Fredrik Johansson.

During the year, F. Johansson has released three new versions (2.10, 2.11 and 2.12) of the Arb software for arbitrary-precision ball arithmetic.

The paper [] describing Arb has been published in the IEEE Transactions on Computers and was selected as the best paper of this journal's Special Issue on Computer Arithmetic. As a result, a video presentation was featured on the journal's website and Johansson was invited to present the paper in a special session at the 24th IEEE Symposium on Computer Arithmetic (ARITH24) at Imperial College London, UK.

In [20], Johansson describes the first complete algorithm for computing the Lambert W function rigorously in complex ball arithmetic.

6.6. Python and Julia computer algebra packages

Participant: Fredrik Johansson.

F. Johansson together with C. Fieker, W. Hart and T. Hofmann of TU Kaiserslautern have developed Nemo and Hecke, two packages for computer algebra and algebraic number theory using the Julia programming language. The paper [16] describing Nemo and Hecke has been published in the proceedings of ISSAC, the main international computer algebra conference.

The paper [14] describing the SymPy package for computer algebra in Python has been published. SymPy is a highly collaborative international project and F. Johansson is one of the 27 coauthors of this paper. Johansson's main contributions to the software include developing the mpmath package used for arbitrary-precision numerical evaluation. In addition, Johansson has issued the stable version 1.0 release of mpmath.

POLSYS Project-Team

6. New Results

6.1. Fundamental algorithms and structured polynomial systems

6.1.1. Linear Algebra for Computing Gröbner Bases of Linear Recursive Multidimensional Sequences

The so-called Berlekamp – Massey – Sakata algorithm computes a Gröbner basis of a 0-dimensional ideal of relations satisfied by an input table. It extends the Berlekamp – Massey algorithm to n-dimensional tables, for $n > 1$.

In [1], we investigate this problem and design several algorithms for computing such a Gröbner basis of an ideal of relations using linear algebra techniques. The first one performs a lot of table queries and is analogous to a change of variables on the ideal of relations.

As each query to the table can be expensive, we design a second algorithm requiring fewer queries, in general. This FGLM-like algorithm allows us to compute the relations of the table by extracting a full rank submatrix of a *multi-Hankel* matrix (a multivariate generalization of Hankel matrices).

Under some additional assumptions, we make a third, adaptive, algorithm and reduce further the number of table queries. Then, we relate the number of queries of this third algorithm to the *geometry* of the final staircase and we show that it is essentially linear in the size of the output when the staircase is convex. As a direct application to this, we decode n-cyclic codes, a generalization in dimension n of Reed Solomon codes.

We show that the multi-Hankel matrices are heavily structured when using the LEX ordering and that we can speed up the computations using fast algorithms for quasi-Hankel matrices. Finally, we design algorithms for computing the generating series of a linear recursive table.

6.1.2. In-depth comparison of the Berlekamp – Massey – Sakata and the Scalar-FGLM algorithms: the non adaptive variants

In [22], we compare thoroughly the BERLEKAMP – MASSEY – SAKATA algorithm and the SCALAR-FGLM algorithm, which compute both the ideal of relations of a multidimensional linear recurrent sequence.

Suprisingly, their behaviors differ. We detail in which way they do and prove that it is not possible to tweak one of the algorithms in order to mimic exactly the behavior of the other.

6.1.3. Resultants and Discriminants for Bivariate Tensor-product Polynomials

Optimal resultant formulas have been systematically constructed mostly for *unmixed polynomial systems*, that is, systems of polynomials which all have the same support. However, such a condition is restrictive, since *mixed systems* of equations arise frequently in practical problems. In [16] we present a square, *Koszul-type* matrix expressing the resultant of arbitrary (mixed) bivariate *tensor-product systems*. The formula generalizes the classical Sylvester matrix of two univariate polynomials, since it expresses a map of *degree one*, that is, the entries of the matrix are simply coefficients of the input polynomials. Interestingly, the matrix expresses a primal-dual multiplication map, that is, the tensor product of a univariate multiplication map with a map expressing derivation in a dual space. Moreover, for tensor-product systems with more than two (affine) variables, we prove an impossibility result: no universal degree-one formulas are possible, unless the system is unmixed. We also present applications of the new construction in the computation of discriminants and mixed discriminants as well as in solving systems of bivariate polynomials with tensor-product structure.

6.1.4. Sparse Rational Univariate Representation

In [15] we present explicit worst case degree and height bounds for the rational univariate representation of the isolated roots of polynomial systems based on mixed volume. We base our estimations on height bounds of resultants and we consider the case of 0-dimensional, positive dimensional, and parametric polynomial systems.

Multi-homogeneous polynomial systems arise in many applications. In [11], we provide bit complexity estimates for representing the solutions of these systems. These are the best currently known bounds. The assumptions essentially imply that the Jacobian matrix of the system under study has maximal rank at the solution set and that this solution set is finite.

We do not only obtain bounds but an algorithm is also given for solving such systems. We give bit complexity estimates which, up to a few extra other factors, are quadratic in the number of solutions and linear in the height of the input system, under some genericity assumptions.

The algorithm is probabilistic and a probability analysis is provided. Next, we apply these results to the problem of optimizing a linear map on the real trace of an algebraic set. Under some genericity assumptions, we provide bit complexity estimates for solving this polynomial minimization problem.

6.1.5. Improving Root Separation Bounds

Let f be a polynomial (or polynomial system) with all simple roots. The root separation of f is the minimum of the pair-wise distances between the complex roots. A root separation bound is a lower bound on the root separation. Finding a root separation bound is a fundamental problem, arising in numerous disciplines. In [7] we present two new root separation bounds: one univariate bound, and one multivariate bound. The new bounds improve on the old bounds in two ways: (1) The new bounds are usually significantly bigger (hence better) than the previous bounds. (2) The new bounds scale correctly, unlike the previous bounds. Crucially, the new bounds are not harder to compute than the previous bounds.

6.1.6. Accelerated Approximation of the Complex Roots and Factors of a Univariate Polynomial

The known algorithms approximate the roots of a complex univariate polynomial in nearly optimal arithmetic and Boolean time. They are, however, quite involved and require a high precision of computing when the degree of the input polynomial is large, which causes numerical stability problems. In [8] we observe that these difficulties do not appear at the initial stages of the algorithms, and in our present paper we extend one of these stages, analyze it, and avoid the cited problems, still achieving the solution within a nearly optimal complexity estimates, provided that some mild initial isolation of the roots of the input polynomial has been ensured. The resulting algorithms promise to be of some practical value for root-finding and can be extended to the problem of polynomial factorization, which is of interest on its own right. We conclude with outlining such an extension, which enables us to cover the cases of isolated multiple roots and root clusters.

6.1.7. Nearly optimal computations with structured matrices

In [9] we estimate the Boolean complexity of multiplication of structured matrices by a vector and the solution of nonsingular linear systems of equations with these matrices. We study four basic and most popular classes, that is, Toeplitz, Hankel, Cauchy and Vandermonde matrices, for which the cited computational problems are equivalent to the task of polynomial multiplication and division and polynomial and rational multipoint evaluation and interpolation. The Boolean cost estimates for the latter problems have been obtained by Kirrinnis, except for rational interpolation. We supply them now as well as the Boolean complexity estimates for the important problems of multiplication of transposed Vandermonde matrix and its inverse by a vector. All known Boolean cost estimates for such problems rely on using Kronecker product. This implies the d-fold precision increase for the d-th degree output, but we avoid such an increase by relying on distinct techniques based on employing FFT. Furthermore we simplify the analysis and make it more transparent by combining the representations of our tasks and algorithms both via structured matrices and via polynomials and rational functions. This also enables further extensions of our estimates to cover Trummer's important

problem and computations with the popular classes of structured matrices that generalize the four cited basic matrix classes, as well as the transposed Vandermonde matrices. It is known that the solution of Toeplitz, Hankel, Cauchy, Vandermonde, and transposed Vandermonde linear systems of equations is generally prone to numerical stability problems, and numerical problems arise even for multiplication of Cauchy, Vandermonde, and transposed Vandermonde matrices by a vector. Thus our FFT-based results on the Boolean complexity of these important computations could be quite interesting because our estimates are reasonable even for more general classes of structured matrices, showing rather moderate growth of the complexity as the input size increases.

6.1.8. Sliding solutions of second-order differential equations with discontinuous right-hand side

In [2], we consider second-order ordinary differential equations with discontinuous right-hand side. We analyze the concept of solution of this kind of equations and determine analytical conditions that are satisfied by typical solutions. Moreover, the existence and uniqueness of solutions and sliding solutions are studied.

6.1.9. Sparse FGLM algorithms

Given a zero-dimensional ideal $I \subset \mathbb{K}[x_1, ..., x_n]$ of degree D, the transformation of the ordering of its Gröbner basis from DRL to LEX is a key step in polynomial system solving and turns out to be the bottleneck of the whole solving process. Thus it is of crucial importance to design efficient algorithms to perform the change of ordering. The main contributions of [3] are several efficient methods for the change of ordering which take advantage of the sparsity of multiplication matrices in the classical *FGLM* algorithm. Combining all these methods, we propose a deterministic top-level algorithm that automatically detects which method to use depending on the input. As a by-product, we have a fast implementation that is able to handle ideals of degree over $40,000$. Such an implementation outperforms the *Magma* and *Singular* ones, as shown by our experiments. First for the shape position case, two methods are designed based on the Wiedemann algorithm: the first is probabilistic and its complexity to complete the change of ordering is $O(D(N_1 + n \log D))$, where N_1 is the number of nonzero entries of a multiplication matrix; the other is deterministic and computes the LEX Gröbner basis of $\sqrt{I}$ via Chinese Remainder Theorem. Then for the general case, the designed method is characterized by the Berlekamp–Massey–Sakata algorithm from Coding Theory to handle the multi-dimensional linearly recurring relations. Complexity analyses of all proposed methods are also provided. Furthermore, for generic polynomial systems, we present an explicit formula for the estimation of the sparsity of one main multiplication matrix, and prove its construction is free. With the asymptotic analysis of such sparsity, we are able to show for generic systems the complexity above becomes $O(\sqrt{6/n\pi} D^{2+\frac{n-1}{n}})$.

6.2. Solving Systems over the Reals and Applications

6.2.1. Answering connectivity queries in real algebraic sets

A roadmap for a semi-algebraic set S is a curve which has a non-empty and connected intersection with all connected components of S. Hence, this kind of object, introduced by Canny, can be used to answer connectivity queries (with applications, for instance, to motion planning) but has also become of central importance in effective real algebraic geometry, since it is used in higher-level algorithms. In [10], we provide a probabilistic algorithm which computes roadmaps for smooth and bounded real algebraic sets. Its output size and running time are polynomial in $(nD)^{n \log d}$, where D is the maximum of the degrees of the input polynomials, d is the dimension of the set under consideration and n is the number of variables. More precisely, the running time of the algorithm is essentially subquadratic in the output size. Even under our assumptions, it is the first roadmap algorithm with output size and running time polynomial in $(nD)^{n \log d}$.

6.2.2. Polynomial optimization and semi-definite programming

In [6], we describe our freely distributed Maple library spectra, for Semidefinite Programming solved Exactly with Computational Tools of Real Algebra. It solves linear matrix inequalities, a fundamental object in effective real algebraic geometry and polynomial optimization, with symbolic computation in exact arithmetic

and it is targeted to small-size, possibly degenerate problems for which symbolic infeasibility or feasibility certificates are required.

The positive semidefinite rank of a convex body C is the size of its smallest positive semi-definite formulation. In [5], we show that the positive semidefinite rank of any convex body C is at least $\sqrt{\log(d)}$ where d is the smallest degree of a polynomial that vanishes on the boundary of the polar of C. This improves on the existing bound which relies on results from quantifier elimination. Our proof relies on the Bézout bound applied to the Karush-Kuhn-Tucker conditions of optimality. We discuss the connection with the algebraic degree of semidefinite programming and show that the bound is tight (up to constant factor) for random spectrahedra of suitable dimension.

6.2.3. The Complexity of an Adaptive Subdivision Method for Approximating Real Curves

In [14] we present the first complexity analysis of the algorithm by Plantinga and Vegter for approximating real implicit curves and surfaces. This approximation algorithm certifies the topological correctness of the output using both subdivision and interval arithmetic. In practice, it has been seen to be quite efficient; our goal is to quantify this efficiency. We focus on the subdivision step (and not the approximation step) of the Plantinga and Vegter algorithm. We begin by extending the subdivision step to arbitrary dimensions. We provide *a priori* worst-case bounds on the complexity of this algorithm both in terms of the number of subregions constructed and the bit complexity for the construction. Then, we use continuous amortization to derive adaptive bounds on the complexity of the subdivided region. We also provide examples showing our bounds are tight.

6.3. Solving Systems in Finite Fields, Applications in Cryptology and Algebraic Number Theory.

6.3.1. Private Multiplication over Finite Fields

The notion of privacy in the probing model, introduced by Ishai, Sahai, and Wagner in 2003, is nowadays frequently involved to assess the security of circuits manipulating sensitive information. However, provable security in this model still comes at the cost of a significant overhead both in terms of arithmetic complexity and randomness complexity. In [13], we deal with this issue for circuits processing multiplication over finite fields. Our contributions are manifold. Extending the work of Belaïd, Benhamouda, Passelègue, Prouff, Thillard, and Vergnaud at Eurocrypt 2016, we introduce an algebraic characterization of the privacy for multiplication in any finite field and we propose a novel algebraic characterization for non-interference (a stronger security notion in this setting). Then, we present two generic constructions of multiplication circuits in finite fields that achieve non-interference in the probing model. The second proposal achieves a linear complexity in terms of randomness consumption. This complexity is proved to be almost optimal. Eventually, we show that our constructions can always be instantiated in large enough finite fields.

6.3.2. Convolutional Neural Networks with Data Augmentation Against Jitter-Based Countermeasures - Profiling Attacks Without Pre-processing

In the context of the security evaluation of cryptographic implementations, profiling attacks (aka Template Attacks) play a fundamental role. Nowadays the most popular Template Attack strategy consists in approximating the information leakages by Gaussian distributions. Nevertheless this approach suffers from the difficulty to deal with both the traces misalignment and the high dimensionality of the data. This forces the attacker to perform critical preprocessing phases, such as the selection of the points of interest and the temporal realignment of measurements. Some software and hardware countermeasures have been conceived exactly to create such a misalignment. In [17], we propose an end-to-end profiling attack strategy based on Deep Learning algorithms combined with Data Augmentation strategies.

6.3.3. Submissions to the NIST Post-Quantum Standardization Process

We have submitted three cryptosystems to the current process leads by `NIST` for standardizing post-quantum public-key algorithms.

6.3.3.1. GeMSS

The acronym stands for a Great Multivariate Signature Scheme [18]. As suggested by its name, *GeMSS* is a multivariate-based signature scheme producing small signatures. It has a fast verification process, and a medium/large public-key. *GeMSS* is in direct lineage from QUARTZ and borrows some design rationale of the Gui multivariate signature scheme. The former schemes are built from the *Hidden Field Equations* crypotsystem (HFE) by using the so-called minus and vinegar modifiers. It is fair to say that HFE and its variants, are the most studied schemes in multivariate cryptography. QUARTZ produces signatures of 128 bits for a security level of 80 bits and was submitted to the *Nessie Ecrypt* competition for public-key signatures. In contrast to many multivariate schemes, no practical attack has been reported against QUARTZ. This is remarkable knowing the intense activity in the cryptanalysis of multivariate schemes.

GeMSS is a faster variant of QUARTZ that incorporates the latest results in multivariate cryptography to reach higher security levels than QUARTZ whilst improving efficiency.

6.3.3.2. DualModeMS

DualModeMS [20] is a multivariate-based signature scheme with a rather peculiar property. Its public-key is small whilst the signature is large. This is in sharp contrast with traditional multivariate signature schemes based on the so-called *Matsumoto and Imai* (MI) principle, such as QUARTZ or Gui, that produce short signatures but have larger public-keys.

DualModeMS is based on the method proposed by A. Szepieniec, W. Beullens, and B. Preneel at PQC'17 where they present a generic technique permitting to transform any (MI-based multivariate signature scheme into a new scheme with much shorter public-key but larger signatures. This technique can be viewed as a *mode of operations* that offers a new flexibility for MI-like signature schemes. Thus, we believe that *DualModeMS* could also be useful for others multivariate-based signature candidates proposed to NIST.

6.3.3.3. CPFKM

CPFKM [19] is a based on the problem of solving a system of noisy non-linear polynomials, also known as the PoSSo with Noise Problem. Our scheme largely borrows its design rationale from key encapsulation schemes based on the Learning With Errors (LWE) problem and its derivatives. The main motivation of building this scheme is to have a key exchange and encapsulation scheme based on the hardness of solving system of noisy polynomials.

6.3.4. The Point Decomposition Problem over Hyperelliptic Curves: toward efficient computations of Discrete Logarithms in even characteristic

Computing discrete logarithms is generically a difficult problem. For divisor class groups of curves defined over extension fields, a variant of the Index-Calculus called Decomposition attack is used, and it can be faster than generic approaches. In this situation, collecting the relations is done by solving multiple instances of the Point m-Decomposition Problem (PDP_m). An instance of this problem can be modelled as a zero-dimensional polynomial system. Solving is done with Gröbner bases algorithms, where the number of solutions of the system is a good indicator for the time complexity of the solving process. For systems arising from a PDP_m context, this number grows exponentially fast with the extension degree. To achieve an efficient harvesting, this number must be reduced as much as possible. Extending the elliptic case, we introduce in [4] a notion of Summation Ideals to describe PDP_m instances over higher genus curves, and compare to Nagao's general approach to PDP_m. In even characteristic we obtain reductions of the number of solutions for both approaches, depending on the curve's equation. In the best cases, for a hyperelliptic curve of genus g, we can divide the number of solutions by $2^{(n-1)(g+1)}$. For instance, for a type II genus 2 curve defined over $\mathbb{F}_{293}$ whose divisor class group has cardinality a near-prime 184 bits integer, the number of solutions is reduced from 4096 to 64. This is enough to build the matrix of relations in around 7 days with 8000 cores using a dedicated implementation.

SECRET Project-Team

7. New Results

7.1. Symmetric cryptology

Participants: Xavier Bonnetain, Christina Boura, Anne Canteaut, Pascale Charpin, Sébastien Duval, Gaëtan Leurent, María Naya Plasencia, Yann Rotella, Ferdinand Sibleyras, Tim Beyne, Mathilde de La Morinerie, André Schrottenloher.

7.1.1. Primitives: block ciphers, stream ciphers, ...

Our recent results mainly concern either the analysis and design of lightweight block ciphers.
Recent results:

- Analysis of linear invariant attacks [41], [54], [28], [29]: C. Beierle, A. Canteaut, G. Leander and Y. Rotella have studied SPN ciphers with a very simple key schedule, such as PRINCE. They introduce properties of the linear layer and of the round constants than can be used to prove that there are no nonlinear invariants.
- Analysis of the probability of differential characteristics for unkeyed constructions [19]: This work shows that the probabilities of some fixed-key differential characteristics are higher than expected when assuming independent S-Boxes. This leads to improved attacks against ROADRUNNER and Minalpher.
- Design and study of a new construction for low-latency block ciphers, named *reflection ciphers*, which generalizes the so-called α-reflection property exploited in PRINCE. This construction aims at reducing the implementation overhead of decryption on top of encryption [15].
- Modular construction of primitives with code-hardness, time-hardness or memory-hardness [42]. A. Biryukov and L. Perrin have introduced new definitions to formalize hardness, and constructions that are hard to compute for common users, but easy for users knowing a secret.
- Design of encryption schemes for efficient homomorphic-ciphertext compression: A. Canteaut, M. Naya-Plasencia together with their coauthors have investigated the constraints on the symmetric cipher imposed by this application and they have proposed some solutions based on additive IV-based stream ciphers [17].

7.1.2. Cryptographic properties and construction of appropriate building blocks

The construction of building blocks which guarantee a high resistance against the known attacks is a major topic within our project-team, for stream ciphers, block ciphers and hash functions. The use of such optimal objects actually leads to some mathematical structures which may be at the origin of new attacks. This work involves fundamental aspects related to discrete mathematics, cryptanalysis and implementation aspects. Actually, characterizing the structures of the building blocks which are optimal regarding to some attacks is very important for finding appropriate constructions and also for determining whether the underlying structure induces some weaknesses or not. For these reasons, we have investigated several families of filtering functions and of S-boxes which are well-suited for their cryptographic properties or for their implementation characteristics.

Recent results:

- Boolean functions with restricted input: Y. Rotella, together with C. Carlet and P. Méaux, has introduced some new criteria on filtering Boolean functions, which measure the security of the recent stream cipher proposal FLIP. Indeed, in this context, the inputs of the filtering function are not uniformly distributed but have a fixed Hamming weight. Then, the main properties of filtering functions (e.g. nonlinearity, algebraic immunity...) have been revisited [20].
- Differential Equivalence of Sboxes: C. Boura, A. Canteaut and their co-authors have studied two notions of differential equivalence of Sboxes corresponding to the case when the functions have the same difference table, or when their difference tables have the same support [45]. They proved that these two notions do not coincide, and that they are invariant under some classical equivalence relations like EA and CCZ equivalence. They also proposed an algorithm for determining the whole equivalence class of a given function.
- A. Canteaut, S. Duval and L. Perrin proposed a construction of a new family of permutations over binary fields of dimension $(4k+2)$ with good cryptographic properties. An interesting property is that this family includes as a specific case the only known APN permutation of an even number of variables [55], [18].
- Construction of cryptographic permutations over finite fields with a sparse representation: P. Charpin, together with N. Cepak and E. Pasalic, exhibited permutations which are derived from sparse functions via linear translators [21].
- New methods for determining the differential spectrum of an Sbox: P. Charpin and G. Kyureghyan have proved that the whole differential spectrum of an Sbox can be determined without examining all derivatives of the mapping, but only the derivatives with respect to an element within a hyperplane [23]. Also, they have proved that, for mappings of a special shape, it is enough to consider the derivatives with respect to all elements within a suitable multiplicative subgroup of $\mathbb{F}_{2^n}$.

7.1.3. Side-channel attacks

Physical attacks must be taken into account in the evaluation of the security of lightweight primitives. Indeed, these primitives are often dedicated to IoT devices in pervasive environments, where an attacker has an easy access to the devices where the primitive is implemented.

Recent results:

- Differential fault attack against LS-designs and SCREAM [52]: this attack generalized previous work on PRIDE to the class of LS-Designs.

7.1.4. Modes of operation and generic attacks

In order to use a block cipher in practice, and to achieve a given security notion, a mode of operation must be used on top of the block cipher. Modes of operation are usually studied through security, and we now that their use is secure as long as the underlying primitive are secure, and we respect some limits on the amount of data processed. The analysis of generic attack helps us understand what happens when the hypothesis of the security proofs do not hold, or the corresponding limits are not respected. Comparing proofs and attack also shows gaps where our analysis is incomplete, and improved proof or attacks are required.

Recent results:

- Use of block ciphers operating on small blocks with the CBC mode [31]: it is well-known that CBC is not secure if the same key is used for encrypting $2^{n/2}$ blocks of plaintext, but this threat has traditionally been dismissed as impractical, even for 64-bit blocks. K. Bhargavan and G. Leurent demonstrated concrete attacks that exploit such short block ciphers in CBC mode.
- Use of block ciphers operating on small blocks with the CTR mode [77]: the security proof of the CTR mode also requires that no more than $2^{n/2}$ blocks are encrypted with the same key, but the known attacks reveal very little information and are considered even less problematic than on CBC. During his internship with G. Leurent, F. Sibleyras has studied concrete attacks against the CTR mode when processing close to $2^{n/2}$ blocks of data, and has shown that an attacker can actually extract as much information as in the case of CBC encryption.

- Improved generic attacks against hash-based MAC [25].
- Modes of operation for full disk encryption [51]: L. Khati, N. Mouha and D. Vergnaud have classified various FDE modes of operation according to their security in a setting where there is no space to store additional data, like an IV or a MAC value. They also introduce the notion of a diversifier, which does not require additional storage, but allows the plaintext of a particular sector to be encrypted into different ciphertexts.

7.2. Code-based cryptography

Participants: Rodolfo Canto Torres, Julia Chaulet, André Chailloux, Thomas Debris, Adrien Hauteville, Nicolas Sendrier, Jean-Pierre Tillich, Matthieu Lequesne, Valentin Vasseur, Matthieu Vieira.

The first cryptosystem based on error-correcting codes was a public-key encryption scheme proposed by McEliece in 1978; a dual variant was proposed in 1986 by Niederreiter. We proposed the first (and only) digital signature scheme in 2001. Those systems enjoy very interesting features (fast encryption/decryption, short signature, good security reduction) but also have their drawbacks (large public key, encryption overhead, expensive signature generation). Some of the main issues in this field are

- security analysis, including against a quantum adversary, implementation and practicality of existing solutions,
- reducing the key size, *e.g.*, by using rank metric instead of Hamming metric, or by using structured codes,
- addressing new functionalities, like identity-based encryption, hashing or symmetric encryption.

As mentioned in Section 5.1.1 , the NIST is currently running a standardization effort for quantum-safe cryptography, where code based cryptography is a promising technique.

Our work in this area can be decomposed as follows:

- suggesting code-based solutions to the NIST competition;
- cryptanalyzing code-based schemes;
- fundamental work on code-based cryptography.

7.2.1. Code-based solutions to the NIST competition

We have proposed two key-exchange protocols to the NIST competition:

- the first one [67] is based on quasi-cyclic MDPC codes and the work [40];
- the second one [69] is based on quasi-cyclic Goppa codes.

Both of them are able to reduce significantly the keysizes by relying on quasi-cyclic codes.

7.2.2. Cryptanalysis of code-based cryptography

Here our work can be summarized as follows:

- cryptanalysis of McEliece schemes based on wild Goppa codes over quadratic extension fields [24];
- improving generic attacks on rank metric codes [68];
- side-channel attacks on quasi-cyclic MDPC bit flipping decoder [74].

7.2.3. Fundamental work on code-based cryptography

- studying precisely the complexity of statistical decoding techniques [71], [48];
- suggesting the first code-based identity-based encryption by using rank metric codes [49];
- suggesting a code-based signature scheme [43];
- analysing and improving the decoding of quasi-cyclic MDPC codes [12], [78];
- studying families of codes that might be used in a cryptographic setting [53].
- improving the complexity of quantum decoding algorithms [50];
- studying [70], [56], [30] whether security reductions for signature schemes are quantum safe when considering the quantum random oracle model (QROM). We were particularly interested in code-based Full Domain Hash constructions. We show that if the underlying correcting code we use has good pseudo random properties then it is possible to perform a quantum security reduction in the QROM.

7.3. Quantum Information

Participants: Xavier Bonnetain, Rémi Bricout, Kaushik Chakraborty, André Chailloux, Shouvik Ghorai, Antoine Grospellier, Anirudh Krishna, Gaëtan Leurent, Anthony Leverrier, Vivien Londe, María Naya Plasencia, Andrea Olivo, Jean-Pierre Tillich, Sristy Agrawal, André Schrottenloher.

Our research in quantum information focusses on several axes: quantum codes with the goal of developing better error correction strategies to build large quantum computers, quantum cryptography which exploits the laws of quantum mechanics to derive security guarantees, relativistic cryptography which exploits in addition the fact that no information can travel faster than the speed of light and finally quantum cryptanalysis which investigates how quantum computers could be harnessed to attack classical cryptosystems.

7.3.1. Quantum codes

Protecting quantum information from external noise is an issue of paramount importance for building a quantum computer. It also worthwhile to notice that all quantum error-correcting code schemes proposed up to now suffer from the very same problem that the first (classical) error-correcting codes had: there are constructions of good quantum codes, but for the best of them it is not known how to decode them in polynomial time.

Recent results:

- Decoding algorithm for quantum expander codes [72], [57], [58], [59], [73], [35]. In this work, A. Grospellier, A. Leverrier and O. Fawzi analyze an efficient decoding algorithm for quantum expander codes and prove that it suppresses errors exponentially in the local stochastic noise model. As an application, this shows that this family of codes can be used to obtain quantum fault-tolerance with only a constant overhead in terms of qubits, compared to a polylogarithmic overhead as in previous schemes. This is a crucial step in order to eventually build large universal quantum computers.
- Construction of quantum LDPC codes from regular tessellations of hyperbolic 4-space [64], [62]. In this work, V. Londe proposes a variant of a construction of Guth and Lubotzky that yields a family of constant rate codes with a polynomial minimum distance. The main interest of this construction is that is is based on a regular tessellation of hyperbolic 4-space by hypercubes. This nice local structure is exploited to design and analyze an efficient decoding algorithm that corrects arbitrary errors of weight logarithmic in the code length.
- Construction of quantum codes based on the real projective space [63]. In this work, V. Londe studies a family of almost LDPC codes with a large minimum distance and another efficient decoding algorithm.
- We were also awarded a European Quantera project “QCDA” to investigate and develop better quantum error-correcting codes and schemes for fault-tolerance.

7.3.2. Quantum cryptography

Quantum cryptography exploits the laws of quantum physics to establish the security of certain cryptographic primitives. The most studied one is certainly quantum key distribution, which allows two distant parties to establish a secret using an untrusted quantum channel. Our activity in this field is particularly focussed on protocols with continuous variables, which are well-suited to implementations. Another primitive is quantum money and was in fact the first proposed idea of quantum cryptography in the 70s. However, this primitive hasn't received much attention because its implementation requires quantum memories, which weren't available until now.

Recent results:

- Full security proof for BB84 [27]. In this work A. Leverrier, with M. Tomamichel, give a detailed and self-contained security proof for BB84, the most studied quantum key distribution protocol. Many simplified proofs appear in the literature, but are usually incomplete and fail to address the whole protocol.
- Security proof of continuous-variable quantum key distribution [26], [36], [37]. In this work, A. Leverrier establishes for the first time a security reduction from general attacks to a class of simple attacks called "collective Gaussian" attacks. This result exploits in a crucial way a recent Gaussian de Finetti theorem that applies to quantum systems of infinite dimension [75], [61], [34].
- In [22], A. Chailloux and I. Kerenidis present an extended version on results for optimal quantum bit commitment and coin flipping. Those results show what is the best way to quantumly perform those protocols in the information-theoretic setting. In the extended version, we also show that the bound for quantum bit commitment cannot be achieved classically, even with an access to an ideal coin flipping primitive.
- We were also awarded an ANR project quBIC and an "Émergence" project from Ville de Paris to study quantum money schemes in collaboration with UPMC, LKB and IRIF.

7.3.3. Relativistic cryptography

Two-party cryptographic tasks are well-known to be impossible without complexity assumptions, either in the classical or the quantum world. Remarkably, such no-go theorems become invalid when adding the physical assumption that no information can travel faster than the speed of light. This additional assumption gives rise to the emerging field of relativistic cryptography. We worked on this topic for several years and Andrea Olivo was recruited as a PhD student to continue working on both theoretical and practical aspects of relativistic cryptography.

Recent results:

- Relativistic zero-knowledge: In [46], A. Chailloux and A. Leverrier construct a relativistic zero-knowledge protocol for any NP complete problem. The main technical tool is the analysis of quantum consecutive measurements, which allows us to prove security against quantum adversaries. While this technique is applied to the relativistic setting, it also has implications for more standard quantum cryptography.
- In [16], R. Bricout and A. Chailloux study relativistic multi-round bit commitment schemes. They show optimal classical cheating strategies for the canonical F_Q commitment scheme. This shows that the security proof derived last year on the relativistic F_Q commitment scheme is essentially optimal against classical adversaries.

7.3.4. Quantum cryptanalysis of symmetric primitives

Symmetric cryptography seems at first sight much less affected in the post-quantum world than asymmetric cryptography: its main known threat seemed for a long time Grover's algorithm, which allows for an exhaustive key search in the square root of the normal complexity. For this reason, it was usually believed that doubling key lengths suffices to maintain an equivalent security in the post-quantum world. However, a lot of work is certainly required in the field of symmetric cryptography in order to "quantize" the classical families of

attacks in an optimized way, as well as to find new dedicated quantum attacks. M. Naya Plasencia has recently been awarded an ERC Starting grant for her project named QUASYModo on this topic, that has started on september 2017.

Recent results:

- In a result published in Asiacrypt 2017 [47] and done during the internship of André Schrottenloher [76] a new quantum algorithm for finding collisions is proposed. The algorithm is based on BHT and exploits distinguished points as well as an improved optimization of the parameters, and allows to find, for the first time, collisions on n bits with a better time complexity than $2^{n/2}$ while needing a polynomial amount of quantum memory.
- Two of the most popular symmetric cryptanalysis families are differential and linear cryptanalysis. In [60] (also presented in [33]), G. Leurent, M. Kaplan, A. Leverrier and M. Naya-Plasencia have proposed efficient ways of quantizing these attacks in different models, obtaining some non-intuitive results: just quantizing the best classical attack does not always provide the best quantum attack.
- X. Bonnetain and M. Naya-Plasencia have obtained some new results, preliminarily described in [14] and presented at [38], that consider the tweak proposed at Eurocrypt this year of using modular additions to counter Simon's attacks. They have studied the best attacks on these constructions, that use Kuperberg's algorithm. They have also simulated the cost of such attacks, improved the algorithm, applied this to a widely-used construction and to some slide attacks, and finally dimensionated the symmetric construction in order to stay secure to these attacks. They have concluded that the proposed tweak does not seam realistic.
- In [44], an attack on the superposition model of the CAESAR cadidate AEZ is proposed, showing that this construction would be completely broken in that scenario.

SPECFUN Project-Team

6. New Results

6.1. Efficient Algorithms in Computer Algebra

This year has seen the end of the writing and the publication of a book on computer-algebra algorithms [8]. The course at Master 2 level *Algorithmes efficaces en calcul formel* is a course that Alin Bostan and Frédéric Chyzak have set up progressively since 2005 together with Marc Giusti (LIX), Bruno Salvy (today AriC), as well as, initially, Éric Schost (LIX at the time) and François Ollivier (LIX), and, more recently, Grégoire Lecerf (LIX). The course is very strongly focused to presenting the design of algorithms guided by complexity analysis, with the goal to lead the students to the understanding of all algorithmic aspects that are necessary to the "creative telescoping" used for symbolic computations of sums and integrals. Their lecture notes had been circulating in and used by the (French) computer-algebra community, while they long had the goal of turning them into a book. They could publish it in 2017 (686 pages), after a big finalization effort in 2016 and 2017. The first parts of the book present fast algorithms for basic objects (integers, polynomials, series, matrices, linear recurrences), insisting on general principles to design efficient algorithms. The next parts of the work build on them to address topics that have made recent progress: factorization of polynomials, algorithms for polynomial systems, definite summation and integration. The work [8] is online as a HAL collection[0]. It is available for free in pdf format and is otherwise sold at a very low price (via print-on-demand). Over the first three months after publication, the book has sold roughly 60 printed copies and the pdf has been downloaded 265 times.

6.2. Hypergeometric Expressions for Generating Functions of Walks with Small Steps in the Quarter Plane

In [2], Alin Bostan and Frédéric Chyzak, together with Mark van Hoeij (Florida State University), Manuel Kauers (Johannes Kepler University), and Lucien Pech, have studied nearest-neighbors walks on the two-dimensional square lattice, that is, models of walks on $\mathbb{Z}^2$ defined by a fixed step set that consists of non-zero vectors with coordinates 0, 1 or -1. They concerned themselves with the enumeration of such walks starting at the origin and constrained to remain in the quarter plane $\mathbb{N}^2$, counted by their length and by the position of their ending point. In earlier works, Bousquet-Mélou and Mishna had identified 19 models of walks that possess a D-finite generating function, and linear differential equations had then been guessed in these cases by Bostan and Kauers. Here, we have given the first proof that these equations are indeed satisfied by the corresponding generating functions. As a first corollary, we have proved that all these 19 generating functions can be expressed in terms of Gauss' hypergeometric functions, with specific parameters that relate them intimately to elliptic integrals. As a second corollary, we have shown that all the 19 generating functions are transcendental, and that among their 19×4 combinatorially meaningful specializations only four are algebraic functions.

6.3. Multiple Binomial Sums

Multiple binomial sums form a large class of multi-indexed sequences, closed under partial summation, which contains most of the sequences obtained by multiple summation of products of binomial coefficients, as well as all the sequences with algebraic generating function. Alin Bostan and Pierre Lairez, together with Bruno Salvy (AriC), have studied in [7] the representation of the generating functions of binomial sums by integrals of rational functions. The outcome is twofold. Firstly, we have shown that a univariate sequence is a multiple binomial sum if and only if its generating function is the diagonal of a rational function. Secondly, we have proposed algorithms that decide the equality of multiple binomial sums and that compute recurrence relations for them. In conjunction with geometric simplifications of the integral representations, this approach behaves well in practice. The process avoids the computation of certificates and the problem of the appearance of spurious singularities that afflicts discrete creative telescoping, both in theory and in practice.

[0] https://hal.archives-ouvertes.fr/AECF/

6.4. Algebraic Diagonals and Walks

The diagonal of a multivariate power series F is the univariate power series $\mathsf{Diag}F$ generated by the diagonal terms of F. Diagonals form an important class of power series; they occur frequently in number theory, theoretical physics and enumerative combinatorics. In [28], Alin Bostan and Louis Dumont, together with Bruno Salvy (AriC), have studied algorithmic questions related to diagonals in the case where F is the Taylor expansion of a bivariate rational function. It is classical that in this case $\mathsf{Diag}F$ is an algebraic function. They have proposed an algorithm for computing an annihilating polynomial of $\mathsf{Diag}F$. They have given a precise bound on the size of this polynomial and show that generically, this polynomial is the minimal polynomial of $\mathsf{Diag}F$ and that its size reaches the bound. Their algorithm runs in time quasi-linear in this bound, which grows exponentially with the degree of the input rational function. They have also addressed the related problem of enumerating directed lattice walks. The insight given by their study has led to a new method for expanding the generating power series of bridges, excursions and meanders. They have shown that their first N terms can be computed in quasi-linear complexity in N, without first computing a very large polynomial equation. An extended version of this work has been presented in [4].

6.5. A Human Proof of the Gessel Conjecture

Counting lattice paths obeying various geometric constraints is a classical topic in combinatorics and probability theory. Many recent works deal with the enumeration of 2-dimensional walks with prescribed steps confined to the positive quadrant. A notoriously difficult case concerns the so-called *Gessel walks*: they are planar walks confined to the positive quarter plane, which move by unit steps in any of the West, North-East, East, and South-West directions. In 2001, Ira Gessel conjectured a closed-form expression for the number of such walks of a given length starting and ending at the origin. In 2008, Kauers, Koutschan and Zeilberger gave a computer-aided proof of this conjecture. The same year, Bostan and Kauers showed, using again computer algebra tools, that the trivariate generating function of Gessel walks is algebraic. This year, Alin Bostan, together with Irina Kurkova (Univ. Paris 6) and Kilian Raschel (CNRS and Univ. Tours), proposed in [6] the first "human proofs" of these results. They are derived from a new expression for the generating function of Gessel walks in terms of special functions.

6.6. Subresultants in Multiple Roots

In [3], we have provided explicit formulae for the coefficients of the order-d polynomial subresultant of $(x-\alpha)^m$ and $(x-\beta)^n$ with respect to the set of Bernstein polynomials $\{(x-\alpha)^j(x-\beta)^{d-j},\ 0 \le j \le d\}$. They are given by hypergeometric expressions arising from determinants of binomial Hankel matrices.

6.7. On Matrices with Displacement Structure: Generalized Operators and Faster Algorithms

For matrices with displacement structure, basic operations like multiplication, inversion, and linear-system solving can all be expressed in terms of a single task: evaluating the product AB, where A is a structured $n \times n$ matrix of displacement rank α, and B is an arbitrary $n \times \alpha$ matrix. Given B and a so-called *generator* of A, this product is classically computed with a cost ranging from $O(\alpha^2\mathsf{M}(n))$ to $O(\alpha^2\mathsf{M}(n)\log(n))$ arithmetic operations, depending on the specific structure of A. (Here, M is a cost function for polynomial multiplication.) In [5], Alin Bostan, jointly with Claude-Pierre Jeannerod (AriC), Christophe Mouilleron (ENSIIE), and Éric Schost (University of Waterloo), has generalized classical displacement operators, based on block diagonal matrices with companion diagonal blocks, and has also designed fast algorithms to perform the task above for this extended class of structured matrices. The cost of these algorithms ranges from $O(\alpha^{\omega-1}\mathsf{M}(n))$ to $O(\alpha^{\omega-1}\mathsf{M}(n)\log(n))$, with ω such that two $n \times n$ matrices over a field can be multiplied using $O(n^\omega)$ field operations. By combining this result with classical randomized regularization techniques, he has obtained faster Las Vegas algorithms for structured inversion and linear system solving.

6.8. Quasilinear Average Complexity for Solving Polynomial Systems

How many operations do we need on the average to compute an approximate root of a random Gaussian polynomial system? Beyond Smale's 17th problem that asked whether a polynomial bound is possible, Pierre Lairez has proved in [10] a quasi-optimal bound (input size)$^{1+o(1)}$, which improves upon the previously known (input size)$^{3/2+o(1)}$ bound. His new algorithm relies on numerical continuation along *rigid continuation paths*. The central idea is to consider rigid motions of the equations rather than line segments in the linear space of all polynomial systems. This leads to a better average condition number and allows for bigger steps. He showed that on the average, one approximate root of a random Gaussian polynomial system of n equations of degree at most D in $n+1$ homogeneous variables can be computed with $O(n^5D^2)$ continuation steps. This is a decisive improvement over previous bounds, which prove no better than $\sqrt{2}^{\min(n,D)}$ continuation steps on the average.

6.9. Computing the Homology of Basic Semialgebraic Sets in Weak Exponential Time

In [9], Pierre Lairez, jointly with Peter Bürgisser (TU Berlin) and Felipe Cucker (City University of Hong Kong), has described and analyzed an algorithm for computing the homology (Betti numbers and torsion coefficients) of basic semialgebraic sets. The algorithm works in weak exponential time, that is, out of a set of exponentially small measure in the space of data, the cost of the algorithm is exponential in the size of the data. All algorithms previously proposed for this problem have a complexity that is doubly exponential (and this is so for almost all data).

6.10. Formally Certified Computation of Improper Definite Integrals

Assia Mahboubi and Thomas Sibut-Pinote, in collaboration with Guillaume Melquiond (Toccata), have pursued their work on the certified computation of intervals approximating the values of definite integrals involving elementary mathematical functions. This library provides an automated tool that builds a formal proof of the correctness of the output, that is, a formal proof that the interval contains the mathematical values and a formal proof of the integrability of the input function on the input interval. This tool has been extended this year, and it can now deal with improper integrals, that is, integrals whose bounds are infinite or singularities of the integrand. The methodology, the implementation and benchmarks have been described in [13].

6.11. A Complete Formal Proof of the Irrationality of $\zeta(3)$

Assia Mahboubi and Thomas Sibut-Pinote have completed a formal proof of the irrationality of the constant $\zeta(3)$. The missing step in a previous work [32] with Frédéric Chyzak and Enrico Tassi was to obtain a formal proof of the asymptotic behaviour of the least common multiple of the first n integers. They have written a report on this work, which is included as a chapter in Thomas Sibut-Pinote's PhD manuscript.

CAIRN Project-Team

7. New Results

7.1. Reconfigurable Architecture Design

7.1.1. Voltage Over-Scaling for Error-Resilient Applications

Participants: Rengarajan Ragavan, Benjamin Barrois, Cédric Killian, Olivier Sentieys.

Voltage scaling has been used as a prominent technique to improve energy efficiency in digital systems, scaling down supply voltage effects in quadratic reduction in energy consumption of the system. Reducing supply voltage induces timing errors in the system that are corrected through additional error detection and correction circuits. In [43], we proposed voltage over-scaling based approximate operators for applications that can tolerate errors. We characterized the basic arithmetic operators using different operating triads (combination of supply voltage, body-biasing scheme and clock frequency) to generate models for approximate operators. Error-resilient applications can be mapped with the generated approximate operator models to achieve optimum trade-off between energy efficiency and error margin. Based on the dynamic speculation technique, best possible operating triad is chosen at runtime based on the user definable error tolerance margin of the application. In our experiments in 28nm FDSOI, we achieved maximum energy efficiency of 89% for basic operators like 8-bit and 16-bit adders at the cost of 20% Bit Error Rate (ratio of faulty bits over total bits) by operating them in near-threshold regime.

7.1.2. Stochastic Computation Elements with Correlated Input Streams

Participants: Rengarajan Ragavan, Rahul Kumar Budhwani, Olivier Sentieys.

In recent years, shrinking size in integrated circuits has imposed a big challenge in maintaining the reliability in conventional computing. Stochastic Computing (SC) has been seen as a reliable, low-cost, and low-power alternative to overcome such issues. SC computes data in the form of bit streams of 1s and 0s. Therefore, SC outperforms conventional computing in terms of tolerance to soft error and uncertainty at the cost of increased computational time. Stochastic Computing with uncorrelated input streams requires streams to be highly independent for better accuracy. This results in more hardware consumption for conversion of binary numbers to stochastic streams. Correlation can be used to design Stochastic Computation Elements (SCE) with correlated input streams. These designs have higher accuracy and less hardware consumption. In [38], we proposed new SC designs to implement image processing algorithms with correlated input streams. Experimental results of proposed SC with correlated input streams show on average 37% improvement in accuracy with reduction of 50-90% in area and 20-85% in delay over existing stochastic designs.

7.1.3. Fault Tolerant Architectures

Participants: Olivier Sentieys, Angeliki Kritikakou, Rafail Psiakis.

Error occurrence in embedded systems has significantly increased, whereas critical applications require reliable processors that combine performance with low cost and energy consumption. Very Long Instruction Word (VLIW) processors have inherent resource redundancy which is not constantly used due to application's fluctuating Instruction Level Parallelism (ILP). Approaches can benefit these additional resources to provide fault tolerance.

The reliability through idle slots utilization can be explored either at compile-time, increasing code size and storage requirements, or at run-time only inside the current instruction bundle, adding unnecessary time slots and degrading performance. To address this issue, we proposed a technique in [41] to explore the idle slots inside and across original and replicated instruction bundles reclaiming more efficiently the idle slots and creating a compact schedule. To achieve this, a dependency analysis is applied at run-time. The execution of both original and replicated instructions is allowed at any adequate function unit, providing higher flexibility on instruction scheduling. The proposed technique achieves up to 26% reduction in performance degradation over existing approaches.

When permanent and soft errors coexist, spare units have to be used or the executed program has to be modified through self-repair or by using several stored versions. However, these solutions introduce high area overhead for the additional resources, time overhead for the execution of the repair algorithm and storage overhead of the multi-versioning. To address these limitations, a hardware mechanism is proposed in [42] which at run-time replicates the instructions and schedules them at the idle slots considering the resource constraints. If a resource becomes faulty, the proposed approach efficiently rebinds both the original and replicated instructions during execution. In this way, the area overhead is reduced, as no spare resources are used, whereas time and storage overhead are not required. Results show up to 49% performance gain over existing techniques.

7.1.4. Hardware Accelerated Simulation of Heterogeneous Platforms

Participants: Minh Thanh Cong, François Charot, Steven Derrien.

When considering designing heterogeneous multi-core platforms, the number of possible design combinations leads to a huge design space, with subtle trade-offs and design interactions. To reason about what design is best for a given target application requires detailed simulation of many different possible solutions. Simulation frameworks exist (such as gem5) and are commonly used to carry out these simulations. Unfortunately, these are purely software-based approaches and they do not allow a real exploration of the design space. Moreover, they do not really support highly heterogeneous multi-core architectures. These limitations motivate the study of the use of hardware to accelerate the simulation, and in particular of FPGA components. In this context, we are currently investigating the possibility of building hardware accelerated simulators using the HAsim simulation infrastructure, jointly developed by MIT and Intel. HAsim is an FPGA-accelerated simulator that is able to simulate a multicore with a high-detailed pipeline, cache hierarchy and detailed on-chip network on a single FPGA. A model of the RISC-V instruction set architecture suited to the HAsim infrastructure has been developed, its deployment on the Xeon+FPGA Intel platform is in progress. This work is done with the perspective of studying hardware accelerated simulation of heterogeneous multicore architectures mixing RISC-V cores and hardware accelerators.

7.1.5. Optical Interconnections for 3D Multiprocessor Architectures

Participants: Jiating Luo, Ashraf El-Antably, Van Dung Pham, Cédric Killian, Daniel Chillet, Olivier Sentieys.

To address the issue of interconnection bottleneck in multiprocessor on a single chip, we study how an Optical Network-on-Chip (ONoC) can leverage 3D technology by stacking a specific photonics die. The objectives of this study target: i) the definition of a generic architecture including both electrical and optical components, ii) the interface between electrical and optical domains, iii) the definition of strategies (communication protocol) to manage this communication medium, and iv) new techniques to manage and reduce the power consumption of optical communications. The first point is required to ensure that electrical and optical components can be used together to define a global architecture. Indeed, optical components are generally larger than electrical components, so a trade-off must be found between the size of optical and electrical parts. For the second point, we study how the interface can be designed to take applications needs into account. From the different possible interface designs, we extract a high-level performance model of optical communications from losses induced by all optical components to efficiently manage Laser parameters. Then, the third point concerns the definition of high-level mechanisms which can handle the allocation of the communication medium for each data transfer between tasks. This part consists in defining the protocol of wavelength allocation. Indeed, the optical wavelengths are a shared resource between all the electrical computing clusters and are allocated at run time according to application needs and quality of service. The last point concerns the definition of techniques allowing to reduce the power consumption of on-chip optical communications. The power of each Laser can be dynamically tuned in the optical/electrical interface at run time for a given targeted bit-error-rate. Due to the relatively high power consumption of such integrated Laser, we study how to define adequate policies able to adapt the laser power to the signal losses.

In [37] we designed an Optical-Network-Interface (ONI) to connect a cluster of several processors to the optical communication medium. This interface, constrained by the 10 Gb/s data-rate of the Lasers, integrates Error Correcting Codes (ECC) and a communication manager. This manager can select, at run-time, the communication mode to use depending on timing or power constraints. Indeed, as the use of ECC is based on redundant bits, it increases the transmission time, but saves power for a given Bit Error Rate (BER). Moreover, our ONI allows for data to be sent using several wavelengths in parallel, hence increasing transmission bandwidth. From the design of this interface, estimation in terms of power consumption and execution time have been obtained, as well as the energy per bit of each communication.

The optical medium can support multiple transactions at the same time on different wavelengths by using Wavelength Division Multiplexing (WDM). Moreover, multiple wavelengths can be gathered as high-bandwidth channel to reduce transmission time. However, multiple signals sharing simultaneously a waveguide lead to inter-channel crosstalk noise. This problem impacts the Signal to Noise Ratio (SNR) of the optical signal, which increases the Bit Error Rate (BER) at the receiver side. In [39], we formulated the crosstalk noise and execution time models and then proposed a Wavelength Allocation (WA) method in a ring-based WDM ONoC to reach performance and energy trade-offs based on the application constraints. We showed that for a 16-core ONoC architecture using 12 wavelengths, more than 10^5 allocation solutions exist and only 51 are on a Pareto front giving a tradeoff between execution time and energy per bit (derived from the BER). These optimized solutions reduce the execution time by 37% or the energy from 7.6fJ/bit to 4.4fJ/bit.

We also proposed to explore the selection of laser power for each communication. This approach reduces the global power consumption by ensuring the targeted Bit Error Rate for each communication. To support laser power selection, we have also studied, designed and evaluated at transistor level different configurable laser drivers using a 28NM FDSOI technology.

7.1.6. Adaptive Dynamic Compilation for Low-Power Embedded Systems

Participants: Steven Derrien, Simon Rokicki.

Single ISA-Heterogeneous multi-cores such as the ARM big.LITTLE have proven to be an attractive solution to explore different energy/performance trade-offs. Such architectures combine Out of Order cores with smaller in-order ones to offer different power/energy profiles. They however do not really exploit the characteristics of workloads (compute-intensive vs. control dominated).

In this work, we propose to enrich these architectures VLIW cores, which are very efficient at compute-intensive kernels. To preserve the single ISA programming model, we resort to Dynamic Binary Translation as used in Transmeta Crusoe and NVidia Denver processors. Our proposed DBT framework targets the RISC-V ISA, for which both OoO and in-order implementations exist.

Since DBT operates at runtime, its execution time is directly perceptible by the user, hence severely constrained. As a matter of fact, this overhead has often been reported to have a huge impact on actual performance, and is considered as being the main weakness of DBT based solutions. This is particularly true when targeting a VLIW processor: the quality of the generated code depends on efficient scheduling; unfortunately scheduling is known to be the most time-consuming component of a JIT compiler or DBT. Improving the responsiveness of such DBT systems is therefore a key research challenge. This is however made very difficult by the lack of open research tools or platform to experiment with such platforms.

To address these issues, we have developed an open hardware/software platform supporting DBT. The platform was designed using HLS tools and validated on a FPGA board. The DBT uses RISC-V as host ISA, and can be retargeted to different VLIW configurations. Our platform uses custom hardware accelerators to improve the reactivity of our optimizing DBT flow. Our results [44] show that, compared to a software implementation, our approach offers speed-up by 8× while consuming 18× less energy.

Our current research work investigates how DBT techniques can be used to support runtime configurable VLIW cores. Such cores enable fine grain exploration of energy/performance trade-off by dynamically adjusting their number of execution slots, their register file size, etc.). More precisely, we build on our DBT framework to enable dynamic code specialization. Our first experimental results suggest that this approach leads to best-case performance and energy efficiency when compared against static VLIW configurations [54].

7.1.7. Design Space Exploration for Iterative Stencil computations on FPGA accelerators

Participants: Steven Derrien, Gaël Deest, Tomofumi Yuki.

Iterative stencil computations arise in many application domains, ranging from medical imaging to numerical simulation. Since they are computationally demanding, a large body of work addressed the problem of parallelizing and optimizing stencils for multi-cores, GPUs, and FPGAs. Earlier attempts targeting FPGAs showed that the performance of such accelerators is the result of a complex interplay between the FPGA's raw computing power, the amount of on-chip memory it has, and the performance of the external memory system. They also illustrate how each application may have different requirements. For example, in the context of embedded vision, the designer's goal is often to find the design with minimum cost that matches real-time performance constraints (e.g., 4K@60fps). In an exascale context, the designer's goal is to maximize performance (measured in ops-per-second) for a given FPGA board, while maintaining power dissipation to a minimum. Based on these observations, we explore a family of design options that can accommodate a large set of requirements and constraints, by exposing trade-offs between computing power, bandwidth requirements, and FPGA resource usage. We have developed a code generator that produces HLS-optimized C/C++ descriptions of accelerator instances targeting emerging System on Chip platforms, (e.g., Xilinx Zynq or Intel SoC). Our family of designs builds upon the well-known tiling transformation, which we use to balance on-chip memory cost and off-chip bandwidth. To ease the exploration of this design space, we propose performance models to hone in on the most interesting design points, and show how they accurately lead to optimal designs. Our results demonstrate that the optimal choice depends on problem sizes and performance goals [30].

7.1.8. Energy-driven Accelerator Exploration for Heterogeneous Multiprocessor Architectures

Participants: Baptiste Roux, Olivier Sentieys.

Programming heterogeneous multiprocessor architectures combining multiple processor cores and hardware accelerators is a real challenge. Computer-aided design and development tools try to reduce the large design space by simplifying hardware software mapping mechanisms. However, energy consumption is not well supported in most of design space exploration methodologies due to the difficulty to fast and accurately estimate energy consumption. To this aim,we proposed and validated an exploration method for partitioning applications on software cores and hardware accelerators under energy-efficiency constraints. The methodology is based on energy and performance measurement of a tiny subset of the design space and an analytical formulation of the performance and energy of an application kernel mapped on a heterogeneous architecture. This closed-form expression is captured and solved using Mixed Integer Linear Programming, which allows for very fast exploration resulting in the optimal solution. The approach is validated on two applications kernels using Zynq-based architecture showing more than 12% acceleration speed-up and energy saving compared to standard approaches. Results also show that the most energy-efficient solution is application- and platform-dependent and moreover hardly predictable, which highlights the need for fast exploration.

7.2. Compilation and Synthesis for Reconfigurable Platform

7.2.1. Superword-Level Parallelism-Aware Word Length Optimization

Participants: Steven Derrien, Ali Hassan El Moussawi.

Many embedded processors do not support floating-point arithmetic in order to comply with strict cost and power consumption constraints. But, they generally provide support for SIMD as a mean to improve performance for little cost overhead. Achieving good performance when targeting such processors requires the use of fixed-point arithmetic and efficient exploitation of SIMD data-path. To reduce time-to-market, automatic SIMDization – such as superword level parallelism (SLP) extraction – and floating-point to fixed-point conversion methodologies have been proposed. In [33], we showed that applying these transformations independently is not efficient. We proposed an SLP-aware word length optimization algorithm to jointly perform floating-point to fixed-point conversion and SLP extraction. We implemented the proposed approach

in a source-to-source compiler framework and evaluated it on several embedded processors. Experimental results illustrated the validity of our approach with performance improvement by up to 40% for a limited loss in accuracy.

7.2.2. Automatic Parallelization Techniques for Time-Critical Systems

Participants: Steven Derrien, Imen Fassi, Thomas Lefeuvre.

Real-time systems are ubiquitous, and many of them play an important role in our daily life. In hard real-time systems, computing the correct results is not the only requirement. In addition, the results must be produced within pre-determined timing constraints, typically deadlines. To obtain strong guarantees on the system temporal behavior, designers must compute upper bounds of the Worst-Case Execution Times (WCET) of the tasks composing the system. WCET analysis is confronted with two challenges: (i) extracting knowledge of the execution flow of an application from its machine code, and (ii) modeling the temporal behavior of the target platform. Multi-core platforms make the latter issue even more challenging, as interference caused by concurrent accesses to shared resources have also to be modeled. Accurate WCET analysis is facilitated by *predictable* hardware architectures. For example, platforms using ScratchPad Memories (SPMs) instead of caches are considered as more predictable. However SPM management is left to the programmer-managed, making them very difficult to use, especially when combined with complex loop transformations needed to enable task level parallelization. Many researches have studied how to combine automatic SPM management with loop parallelization at the compiler level.It has been shown that impressive average-case performance improvements could be obtained on compute intensive kernels, but their ability to reduce WCET estimates remains to be demonstrated, as the transformed code does not lends itself well to WCET analysis.

In the context of the ARGO project, and in collaboration with members of the PACAP team, we have studied how parallelizing compilers techniques should be revisited in order to help WCET analysis tools. More precisely, we have demonstrated the ability of polyhedral optimization techniques to reduce WCET estimates in the case of sequential codes, with a focus on locality improvement and array contraction. We have shown on representative real-time image processing use cases that they could bring significant improvements of WCET estimates (up to 40%) provided that the WCET analysis process is guided with automatically generated flow annotations [31].

7.2.3. Operator-Level Approximate Computing

Participants: Benjamin Barrois, Olivier Sentieys.

Many applications are error-resilient, allowing for the introduction of approximations in the calculations, as long as a certain accuracy target is met. Traditionally, fixed-point arithmetic is used to relax accuracy, by optimizing the bit-width. This arithmetic leads to important benefits in terms of delay, power and area. Lately, several hardware approximate operators were invented, seeking the same performance benefits. However, a fair comparison between the usage of this new class of operators and classical fixed-point arithmetic with careful truncation or rounding, has never been performed. In [27], we first compare approximate and fixed-point arithmetic operators in terms of power, area and delay, as well as in terms of induced error, using many state-of-the-art metrics and by emphasizing the issue of data sizing. To perform this analysis, we developed a design exploration framework, *ApxPerf*, which guarantees that all operators are compared using the same operating conditions. Moreover, operators are compared in several classical real-life applications leveraging relevant metrics. In [27], we show that considering a large set of parameters, existing approximate adders and multipliers tend to be dominated by truncated or rounded fixed-point ones. For a given accuracy level and when considering the whole computation data-path, fixed-point operators are several orders of magnitude more accurate while spending less energy to execute the application. A conclusion of this study is that the entropy of careful sizing is always lower than approximate operators, since it require significantly less bits to be processed in the data-path and stored. Approximated data therefore always contain on average a greater amount of costly erroneous, useless information.

In [26] we performed a comparison between custom fixed-point (FxP) and floating-point (FlP) arithmetic, applied to bidimensional K-means clustering algorithm. First, FxP and FlP arithmetic operators are compared in terms of area, delay and energy, for different bitwidth, using the *ApxPerf2.0* framework. Finally, both are compared in the context of K-means clustering. The direct comparison shows the large difference between 8-to-16-bit FxP and FlP operators, FlP adders consuming 5-12$\times$ more energy than FxP adders, and multipliers 2-10$\times$ more. However, when applied to K-means clustering algorithm, the gap between FxP and FlP tightens. Indeed, the accuracy improvements brought by FlP make the computation more accurate and lead to an accuracy equivalent to FxP with less iterations of the algorithm, proportionally reducing the global energy spent. The 8-bit version of the algorithm becomes more profitable using FlP, which is 80% more accurate with only $1.6\times$ more energy.

7.2.4. Dynamic Fault-Tolerant Mapping and Scheduling on Multi-core systems

Participants: Emmanuel Casseau, Petr Dobias.

Demand on multi-processor systems for high performance and low energy consumption still increases in order to satisfy our requirements to perform more and more complex computations. Moreover, the transistor size gets smaller and their operating voltage is lower, which goes hand in glove with higher susceptibility to system failure. In order to ensure system functionality, it is necessary to conceive fault-tolerant systems. One way to tackle this issue is to makes use of both the redundancy and reconfigurable computing, especially when multi-processor platforms are targeted. Actually, multi-processor platforms can be less vulnerable when one processor is faulty because other processors can take over its scheduled tasks.

In this context, we investigate how to dynamically map and schedule tasks onto homogeneous faulty processors. We developed a run-time algorithm based on the primary/backup approach which is commonly used for its minimal resources utilization and high reliability. Its principal rule is that, when a task arrives, the system creates two identical copies: the primary copy and the backup copy. Several policies have been studied and their performances have been analyzed. We are currently refining the algorithm to reduce its complexity without decreasing performance. This work is done in collaboration with Oliver Sinnen, PARC Lab., the University of Auckland.

7.2.5. Energy Constrained and Real-Time Scheduling and Mapping on Multicores

Participants: Olivier Sentieys, Angeliki Kritikakou, Lei Mo.

Multicore architectures are now widely used in energy-constrained real-time systems, such as energy-harvesting wireless sensor networks. To take advantage of these multicores, there is a strong need to balance system energy, performance and Quality-of-Service (QoS). The Imprecise Computation (IC) model splits a task into mandatory and optional parts allowing to tradeoff QoS. We focus on the problem of mapping, i.e. allocating and scheduling, IC-tasks to a set of processors to maximize system QoS under real-time and energy constraints, which we formulate as a Mixed Integer Linear Programming (MILP) problem. However, state-of-the-art solving techniques either demand high complexity or can only achieve feasible (suboptimal) solutions. We develop an effective decomposition-based approach in [40] to achieve an optimal solution while reducing computational complexity. It decomposes the original problem into two smaller easier-to-solve problems: a master problem for IC-tasks allocation and a slave problem for IC-tasks scheduling. We also provide comprehensive optimality analysis for the proposed method. Through the simulations, we validate and demonstrate the performance of the proposed method, resulting in an average 55% QoS improvement with regards to published techniques.

7.2.6. Real-Time Scheduling of Reconfigurable Battery-Powered Multi-Core Platforms

Participants: Daniel Chillet, Aymen Gammoudi.

Reconfigurable real-time embedded systems are constantly increasingly used in applications like autonomous robots or sensor networks. Since they are powered by batteries, these systems have to be energy-aware, to adapt to their environment and to satisfy real-time constraints. For energy harvesting systems, regular recharges of battery can be estimated, and by including this parameter in the operating system, it is then possible to develop strategy able to ensure the best execution of the application until the next recharge. In this context, operating system services must control the execution of tasks to meet the application constraints. Our objective concerns the proposition of a new real-time scheduling strategy that considers execution constraints such as the deadline of tasks and the energy for heterogeneous architectures. For such systems, we first addressed homogeneous architectures and extended our work for heterogeneous systems for which each task has different execution parameters. For these two architectures models, we formulated the problem as an ILP optimisation problem that can be solved by classical solvers. Assuming that the energy consumed by the communication is dependent on the distance between processors, we proposed a mapping strategy to minimise the total cost of communication between processors by placing the dependent tasks as close as possible to each other. The proposed strategy guarantees that, when a task is mapped into the system and accepted, it is then correctly executed prior to the task deadline. Finally, as on-line scheduling is targeted for this work, we proposed heuristics to solve these problems in efficient way. These heuristics are based on the previous packing strategy developed for the mono-processor architecture case.

7.2.7. Run-Time Management on Multicore Platforms

Participant: Angeliki Kritikakou.

In real-time mixed-critical systems, Worst-Case Execution Time analysis (WCET) is required to guarantee that timing constraints are respected —at least for high criticality tasks. However, the WCET is pessimistic compared to the real execution time, especially for multicore platforms. As WCET computation considers the worst-case scenario, it means that whenever a high criticality task accesses a shared resource in multi-core platforms, it is considered that all cores use the same resource concurrently. This pessimism in WCET computation leads to a dramatic under utilization of the platform resources, or even failing to meet the timing constraints. In order to increase resource utilization while guaranteeing real-time guarantees for high criticality tasks, previous works proposed a run-time control system to monitor and decide when the interferences from low criticality tasks cannot be further tolerated. However, in the initial approaches, the points where the controller is executed were statically predefined. We propose a dynamic run-time control in [19] which adapts its observations to on-line temporal properties, increasing further the dynamism of the approach, and mitigating the unnecessary overhead implied by existing static approaches. Our dynamic adaptive approach allows to control the ongoing execution of tasks based on run-time information, and increases further the gains in terms of resource utilization compared with static approaches.

CAMUS Team

7. New Results

7.1. Automatic (Un-)Collapsing of Non-Rectangular Loops

Participants: Philippe Clauss, Ervin Altıntaş, Matthieu Kuhn.

Loop collapsing is a well-known loop transformation which combines some loops that are perfectly nested into one single loop. It allows to take advantage of the whole amount of parallelism exhibited by the collapsed loops, and provides a perfect load balancing of iterations among the parallel threads.

However, in the current implementations of this loop optimization, as the ones of the OpenMP language, automatic loop collapsing is limited to loops with constant loop bounds that define rectangular iteration spaces, although load imbalance is a particularly crucial issue with non-rectangular loops. The OpenMP language addresses load balance mostly through dynamic runtime scheduling of the parallel threads. Nevertheless, this runtime schedule introduces some unavoidable execution-time overhead, while preventing to exploit the entire parallelism of all the parallel loops.

We propose a technique to automatically collapse any perfectly nested loops defining non-rectangular iteration spaces, whose bounds are linear functions of the loop iterators. Such spaces may be triangular, tetrahedral, trapezoidal, rhomboidal or parallelepiped. Our solution is based on original mathematical results addressing the inversion of a multi-variate polynomial that defines a ranking of the integer points contained in a convex polyhedron.

We show on a set of non-rectangular loop nests that our technique allows to generate parallel OpenMP codes that outperform the original parallel loop nests, parallelized either by using options “static” or “dynamic” of the OpenMP-schedule clause. A conference paper presenting these results, co-authored by Philippe Clauss, Ervin Altıntaş (Master student) and Matthieu Kuhn (Inria Bordeaux Sud-Ouest, team HIEPACS), has been published at the International Parallel and Distributed Processing Symposium (IPDPS) [15].

We are currently developing a technique to also provide good load balancing when parallelizing non-rectangular loops carrying dependences. This new technique has been called *loop uncollapsing*. The idea is to split the outermost parallel loop into two nested loops, such that the new outermost loop, when parallelized, results in well-balanced parallel threads.

7.2. Code-Bones for Fast and Flexible Runtime Code Generation

Participants: Juan Manuel Martinez Caamaño, Manuel Selva, Philippe Clauss.

We have developed a new runtime code generation technique for speculative loop optimization and parallelization. The main benefit of this technique, compared to previous approaches, is to enable advanced optimizing loop transformations at runtime with an acceptable time overhead. The loop transformations that may be applied are those handled by the polyhedral model. The proposed code generation strategy is based on the generation of *code-bones* at compile-time, which are parametrized code snippets either dedicated to speculation management or to computations of the original target program. These code bones are then instantiated and assembled at runtime to constitute the speculatively-optimized code, as soon as an optimizing polyhedral transformation has been determined. Their granularity threshold is sufficient to apply any polyhedral transformation, while still enabling fast runtime code generation. This approach has been implemented in the speculative loop parallelizing framework Apollo, and has been more recently extended to also support loops exhibiting a non-linear behavior thanks to a modeling using “tubes”. The whole approach has been published in Concurrency and Computation: Practice and Experience [11].

7.3. Formal Proofs about Explicitly Parallel Programs with Clocks

Participants: Alain Ketterlin, Éric Violard, Tomofumi Yuki, Paul Feautrier.

We have continued this year our work on formalizing the *happens-before* relation in explicitly parallel programs of the X10 family. Our goal is to define, for certain classes of programs, a relation between instances of elementary instructions that guarantees that one instance necessarily executes before another. Our toy language includes static-control counted loops and conditionals, as well as the usual `finish` and `async` parallel constructs. Moreover, parallel activities can synchronize though the use of *clocks*, which are barriers with dynamic membership. Clocks partition the execution into phases, and profoundly modify the happens-before relation.

This year's work has focused on correctly accounting for the possibility to define specific activities that execute irrespective of the discipline of the clock in scope, so-called *detached* activities. The presence of such activities modifies the notion of phase number, because they let their instructions execute across a range of clock-phases. Our generic notion of phase *ranking* had to be modified. Similarly, the natural semantics we defined had to be slightly modified to correctly represent the parallel execution of both clocked and detached activities. In practice, almost every lemma of the Coq proof has changed, and new definitions were introduced. The new definition of happens-before preserves all desirable properties: it is correct and complete, and is a strict partial order. There is one unpleasant aspect of detached activities that had a strong impact on happens-before: the possibility of deadlocks. A significant part of new definitions and lemmas are devoted to explicit the conditions under which programs terminate. A useful outcome of this part of the mechanization is a static, compile-time deadlock detection criterion.

Most of this work has been described in a paper currently under submission (this paper will be on HAL as soon as anonymity constraints permit). However, the diversity of themes covered in this research (compilation of static-control programs, especially those that fit the polyhedral model, but also semantic modeling of explicitly parallel programs, and formal proofs) make us contemplate the redaction of a much longer paper, which we plan to start at the beginning of next year. At the same time, this work (especially the part about deadlocks) led us to start designing an happens-before relation for a language where multiple clocks can share (part of) their scopes. We hope to be able to advance the formalization of this new family of languages in the near future.

7.4. High-Performance Particle-in-Cell Simulations

Participants: Arthur Charguéraud, Yann Barsamian, Alain Ketterlin.

Yann Barsamian's PhD thesis focuses on the development of efficient programs for Particle-in-Cell (PIC) simulations, with application to plasma physics. Typically, a simulation involves a cluster of machines, each machine hosting several cores, and each core being able to execute vectorized instructions (SIMD). The challenge is to efficiently exploit these three levels of parallelism. Regarding the processing on one given multicore machine, existing algorithms either suffer from suboptimal execution time, due to sorting operations or use of atomic instructions, or suffer from suboptimal space usage. We have developed a novel parallel algorithm for PIC simulations on multicore hardware that features asymptotically-optimal memory consumption, and that does not perform unnecessary accesses to the main memory. The algorithm relies on the use of *chunk bags*, i.e., linked lists of fixed-capacity arrays, for storing particles and allowing to process them efficiently using SIMD instructions. Practical results show excellent scalability on the classical Landau damping and two-stream instability test cases. A paper was published at PPAM [12].

7.5. Granularity Control for Parallel Programs

Participant: Arthur Charguéraud.

Arthur Charguéraud contributes to the ERC DeepSea project, which is hosted at Inria Paris (team Gallium). With his co-authors, he focused this year on the development of techniques for controlling granularity in parallel programs. Granularity control is an essential problem because creating too many tasks may induce overwhelming overheads, while creating too few tasks may harm the ability to process tasks in parallel. Granularity control turns out to be especially challenging for nested parallel programs, i.e., programs in which parallel constructs such as fork-join or parallel-loops can be nested arbitrarily. This year, the DeepSea team investigated two different approaches.

The first one is based on the use of asymptotic complexity functions provided by the programmer, combined with runtime measurements to estimate the constant factors that apply. Combining these two sources of information allows to predict with reasonable accuracy the execution time of tasks. Such predictions may be used to guide the generation of tasks, by sequentializing computations of sufficiently-small size. An analysis is developed, establishing that task creation overheads are indeed bounded to a small fraction of the total runtime. These results extend prior work by the same authors [29], extending them with a carefully-designed algorithm for ensuring convergence of the estimation of the constant factors deduced from the measures, even in the face of noise and cache effects, which are taken into account in the analysis. The approach is demonstrated on a range of benchmarks taken from the state-of-the-art PBBS benchmark suite. These results were submitted to an international conference.

The second approach is based on an instrumentation of the runtime system. The idea is to process parallel function calls just like normal function calls, by pushing a frame on the stack, and only subsequently promoting these frames as threads that might get scheduled on other cores. The promotion of frames takes place at regular time interval, hence the name *heartbeat scheduling* given to the approach. Unlike in prior approaches such as *lazy scheduling*, in which promotion is guided by the work load of the system, hearbeat scheduling can be proved to induce only small scheduling overheads, and to not reduce asymptotically the amount of parallelism inherent to the parallel program. The theory behind the approach is formalized in Coq. It is also implemented through instrumented C++ programs, and evaluated on PBBS benchmarks. A paper describing this approach was submitted to an international conference.

7.6. Program verification and formal languages

Participant: Arthur Charguéraud.

- A. Charguéraud and François Pottier (Inria Paris) extended their formalization of the correctness and asymptotic complexity of the classic Union Find data structure, which features the bound expressed in terms of the inverse Ackermann function. The proof, conducted using CFML extended with time credits, was refined using a slightly more complex potential function, allowing to derive a simpler and richer interface for the data structure. This work appeared in the Journal of Automated Reasoning (JAR) [9].
- A. Charguéraud and F. Pottier have developed an extension of Separation Logic with temporary read-only permissions. This mechanism allows to temporarily convert any assertion (or "permission") to a read-only form. Unlike with fractional permissions, no accounting is required: the proposed read-only permissions can be freely duplicated and discarded. Where mutable data structures are temporarily accessed only for reading, the proposed read-only permissions enable more concise specifications and proofs. All the metatheory is verified in Coq. An article was presented at ESOP [14].
- Armaël Guéneau, PhD student advised by A. Charguéraud and F. Pottier, has developed a Coq library formalizing the asymptotic notation (big-O), and has developed an extension of the CFML verification tool to allow specifying the asymptotic complexity of higher-order, imperative programs. This new feature has been tested on several classic examples of complexity analyses, including: nested loops in $O(n^3)$ and $O(nm)$, selection sort in $O(n^2)$, recursive functions in $O(n)$ and $O(2^n)$, binary search in $O(\log n)$, and Union-Find in $O(\alpha(n))$. A paper was submitted paper to an international conference.
- A. Charguéraud has made progress towards CFML 2.0, a reimplementation of CFML entirely inside Coq. In contrast, the initial version of CFML, developed in A. Charguéraud's PhD thesis, is based on an external tool that parses OCaml source code and produces Coq axioms describing their semantics. The new version will remove the need for axioms, thereby further reducing the trusted code base. Furthermore, CFML 2.0 provides a more general memory model, designed to also accomodate formal reasoning about C-style programs, in future work. In passing, A. Charguéraud performed a complete cleanup of the TLC Coq library, which is used extensively by CFML, leading to the beta release of TLC 2.0.

- A. Charguéraud, together with Alan Schmitt (Inria Rennes) and Thomas Wood (Imperial College), developed an interactive debugger for JavaScript. The interface, accessible as a webpage in a browser, allows to execute a given JavaScript program, following step by step the formal specification of JavaScript developed in prior work on *JsCert* [31]. Concretely, the tool acts as a double-debugger: one can visualize both the state of the interpreted program and the state of the interpreter program. This tool is intended for the JavaScript committee, VM developers, and other experts in JavaScript semantics. A paper describing the tool has been submitted, and the tool has been presented to the JavaScript standardization committee (ECMA) in November 2017.

7.7. Combining Locking and Data Management Interfaces

Participants: Jens Gustedt, Mariem Saied, Daniel Salas.

Handling data consistency in parallel and distributed settings is a challenging task, in particular if we want to allow for an easy to handle asynchronism between tasks. Our publication [1] shows how to produce deadlock-free iterative programs that implement strong overlapping between communication, IO and computation.

An implementation (ORWL) of our ideas of combining control and data management in C has been undertaken, see Section 6.8 . In previous work it has demonstrated its efficiency for a large variety of platforms.

This year, wee have been able to use the knowledge of the communication structure of ORWL programs to map tasks to cores and thereby achieve interesting performance gains on multicore architectures, see [16]. We propose a topology-aware placement module that is based on the Hardware Locality framework, HWLOC, and that takes the characteristics of the application, of the runtime and of the architecture into account. The aim is double. On one hand we increase the abstraction and the portability of the framework, and on the other hand we enhance the performance of the model's runtime.

Within the framework of the thesis of Daniel Salas we have successfully applied ORWL to process large histopathology images. We are now able to treat such images distributed on several machines or shared in an accelerator (Xeon Phi) transparently for the user.

7.8. Automatic Generation of Adaptive Simulation Codes

Participants: Cédric Bastoul, Maxime Schmitt.

Compiler automatic optimization and parallelization techniques are well suited for some classes of simulation or signal processing applications, however they usually don't take into account neither domain-specific knowledge nor the possibility to change or to remove some computations to achieve "good enough" results. Quite differently, production simulation and signal processing codes have adaptive capabilities: they are designed to compute precise results only where it matters if the complete problem is not tractable or if the computation time must be short. In this research, we design a new way to provide adaptive capabilities to compute-intensive codes automatically, inspired by Adaptive Mesh Refinement a classical numerical analysis technique to achieve precise computation only in pertinent areas. It relies on domain-specific knowledge provided through special pragmas by the programmer in the input code and on polyhedral compilation techniques, to continuously regenerate at runtime a code that performs heavy computations only where it matters at every moment. A case study on a fluid simulation application shows that our strategy enables dramatic computation savings in the optimized portion of the application while maintaining good precision, with a minimal effort from the programmer.

This research direction started in 2015 and complements our other efforts on dynamic optimization. In 2016, we started a collaboration on this topic with Inria Nancy - Grand Est team TONUS, specialized on applied mathematics (contact: Philippe Helluy), to bring models and techniques from this field to compilers. This collaboration received the support from the excellence laboratory (LabEx) IRMIA through the funding of the thesis of Maxime Schmitt on this topic. Two papers on this new research direction has been accepted this year on this topic (IMPACT 2017 workshop, HiPC 2017 conference [20]).

7.9. Parallel Polyhedral Regions

Participants: Cédric Bastoul, Vincent Loechner, Harenome Ranaivoarivony-Razanajato.

Nowadays best performing automatic parallelizers and data locality optimizers for static control programs rely on the polyhedral model. State-of-the-art polyhedral compilers generate only one type of parallelism when targeting multicore shared memory architectures: parallel loops via the OpenMP `omp parallel for` directive.

We propose to explore how a polyhedral compiler could exploit parallel region constructs. Instead of initializing a new set of threads each time the code enters a parallel loop and synchronizing them when exiting it, the threads are initialized once for all at the entrance of the region of interest, and synchronized only when it is necessary.

Technically, the whole region containing parallel loops is embedded in an `omp parallel` construct. Inside the parallel region, the `single` construct is used when some code needs to be executed sequentially; the `for` construct is used to distribute loop iterations between threads. Thanks to the power of the polyhedral dependence analysis, we compute when it is valid to add the optional `nowait` clause, to omit the implicit barrier at the end of a worksharing construct and thus to reduce even more control overhead.

This work was published and presented at the HiPC 2017 conference [19].

7.10. Optimization of Sparse Triangular and Banded Matrix Codes

Participants: Vincent Loechner, Rachid Seghir, Toufik Baroudi.

This work is a collaboration between Vincent Loechner and Rachid Seghir from University of Batna (Algeria). Toufik Baroudi is a second year PhD student under his supervision. Rachid Seghir was visiting the CAMUS team from March 25th to April 8th, 2017.

In this work, we enabled static polyhedral optimization techniques to handle sparse matrix storage formats. When handling sparse triangular and banded matrices in their packed formats, such as in the LAPACK library band storage, loop nests bounds and array references of the resulting codes are not affine functions. We proposed to use a new 2d-packed layout and simple affine transformations to enable polyhedral optimization of sparse triangular and banded matrix operations. The effectiveness of our proposal was shown through an experimental study over a large set of linear algebra benchmarks.

These results were published in ACM TACO [8], and will be presented at the HiPEAC conference in January 2018.

CORSE Project-Team

6. New Results

6.1. Simplification and Run-time Resolution of Data Dependence Constraints for Loop Transformations

Participants: Diogo Nunes Sampaio, Alain Ketterlin [Inria CAMUS], Louis-Noël Pouchet [CSU, USA], Fabrice Rastello.

Loop optimizations such as tiling, thread-level parallelization or vectorization are essential transformations to improve performance. Their use rely on the ability to compute dependence information at compile-time to assess their validity, but in many real situations, dependence analysis fails to provide precise enough information. Typical examples where this happens are when working over compilers IR (e.g., LLVM IR) or with legacy source code, with pointers and linearized arrays (e.g., packed symmetric matrices in BLAS LAPACK). In this scenario, the compiler will often be unable to apply aggressive transformations due to lack of conclusive static dependence analysis.

This work makes a fundamental leap towards enabling complex loop transformations in real-life scenarios, by using a hybrid static+dynamic analysis to disambiguate may-dependencies. Similarly to GCC's auto-vectorization, our approach consists in adding a lightweight run-time test to check whether ambiguous may-dependencies do exist at execution time, to determine whether the optimized or unmodified code version should be called. The main contribution of our work is to generalize this pragmatic approach to a large class of loop-nest transformations, including tiling, loop invariant code motion, parallelization, etc. In particular, we design a quantifier elimination scheme on integer multivariate-polynomials, which can aid application of off-the-shelf polyhedral transformations on a larger class of programs, that holds polynomial memory access and affine loop bounds.

The preciseness of the presented scheme and the low run-time overhead of the test are key to make this approach realistic. We experimentally validate our technique on 25 benchmarks using complex loop transformations, achieving negligible overhead. Preciseness is assessed by the observed success of generated test in practical cases.

IPFME tool 5.5 has been developped in this context. This work is the fruit of the collaboration 8.4.1.1 with OSU. It has been presented at the ACM/SIGARCH International Conference on Supercomputing, ICS 2017 [25].

6.2. Optimizing the Four-Index Integral Transform Using Data Movement Lower Bounds Analysis

Participants: Samyam Rajbhandari [Microsoft, USA], Fabrice Rastello, Karol Kowalski [PNNL, USA], Sriram Krishnamoorthy [PNNL, USA], P. Sadayappan [OSU, USA].

The four-index integral transform is a fundamental and computationally demanding calculation used in many computational chemistry suites such as NWChem. It transforms a four-dimensional tensor from an atomic basis to a molecular basis. This transformation is most efficiently implemented as a sequence of four tensor contractions that each contract a four-dimensional tensor with a two-dimensional transformation matrix. Differing degrees of permutation symmetry in the intermediate and final tensors in the sequence of contractions cause intermediate tensors to be much larger than the final tensor and limit the number of electronic states in the modeled systems.

Loop fusion, in conjunction with tiling, can be very effective in reducing the total space requirement, as well as data movement. However, the large number of possible choices for loop fusion and tiling, and data/computation distribution across a parallel system, make it challenging to develop an optimized parallel implementation for the four-index integral transform. We develop a novel approach to address this problem, using lower bounds modeling of data movement complexity. We establish relationships between available aggregate physical memory in a parallel computer system and ineffective fusion configurations, enabling their pruning and consequent identification of effective choices and a characterization of optimality criteria. This work has resulted in the development of a significantly improved implementation of the four-index transform that enables higher performance and the ability to model larger electronic systems than the current implementation in the NWChem quantum chemistry software suite.

This work is the fruit of the collaboration 8.4.1.1 with OSU. It has been presented at the ACM/SIGPLAN Symposium on Principles and Practice of Parallel Programming, PPoPP 2017 [21].

6.3. Register Optimizations for Stencils on GPUs

Participants: Aravind Sukumaran-Rajam [OSU, USA], Atanas Rountev [OSU, USA], Fabrice Rastello, Louis-Noël Pouchet [CSU, USA], P. Sadayappan [OSU, USA].

The recent advent of compute-intensive GPU architecture has allowed application developers to explore high-order 3D stencils for better computational accuracy. A common optimization strategy for such stencils is to expose sufficient data reuse by means such as loop unrolling, with the hope of register-level reuse. However, the resulting code is often highly constrained by register pressure. While the current state-of-the-art register allocators are satisfactory for most applications, they are unable to effectively manage register pressure for such complex high-order stencils, resulting in a sub-optimal code with a large number of register spills. In this paper, we develop a statement reordering framework that models stencil computations as DAG of trees with shared leaves, and adapts an optimal scheduling algorithm for minimizing register usage for expression trees. The effectiveness of the approach is demonstrated through experimental results on a range of stencils extracted from application codes.

This work is the fruit of the collaboration 8.4.1.1 with OSU. It will be presented at the ACM/SIGPLAN Symposium on Principles and Practice of Parallel Programming, PPoPP 2018.

6.4. Data-Flow/Dependence Profiling for Structured Transformations

Participants: Diogo Nunes Sampaio, Fabian Gruber, Christophe Guillon [STMicroelectronics], Antoine Moynault [STMicroelectronics], Louis-Noël Pouchet [CSU, USA], Fabrice Rastello.

Profiling feedback is an important technique used by developers for performance debugging, where it is usually used to pinpoint performance bottlenecks and also to find optimization opportunities. Assessing the validity and potential benefit of a program transformation requires accurate knowledge of the data flow and data dependencies, which can be uncovered by profiling a particular execution of the program.

In this work we develop MICKEY 5.4 , an end-to-end infrastructure for dynamic binary analysis, which produces feedback about the potential to apply structured transformations to uncover non-trivial parallelism and data locality via complex program re-scheduling. Our tool can handle both inter- and intra-procedural aspects of the program in a unified way, thus enabling inter-procedural structured transformations. It is based on QEMU and uses dynamic binary translation to instrument arbitrary programs at run-time. The design of this tool was driven by the goal of achieving portability, both in terms of targeted CPU architectures, but also in terms of programming environment and the use of third-party libraries for which no source code is available.

This work is the fruit of the collaboration 8.4.1.1 with CSU and the contract 7.2 with STMicroelectronics.

6.5. Dynamic Load Balancing of Monte Carlo Particle Transport Applications

Participants: Thomas Gonçalves, Marc Pérache [CEA DAM, Bruyères le Châtel], Frederic Desprez, Jean-Francois Mehaut.

Monte Carlo particle transport applications consist in studying the behavior of particles moving about a simulation domain. Particle distribution among simulation domains is not uniform and changes dynamically during simulation. The parallelization of this kind of applications on massively parallel architectures leads to solve complex issues of workloads and data balancing among numerous compute cores.

This research work started by identifying parallelization pitfalls of Monte Carlo particle transport applications using theoretical and experimental analysis of reference parallelization methods. A semi-dynamic load-balancing based on partitioning techniques has then been proposed. Finally, we designed a dynamic approach which is able to redistribute workloads and data while keeping a low communication volume. Compared to the perfectly balanced domain replication method using strong scaling measurement, the dynamic approach leads both to speedups and reduction of memory footprint.

This work is part of the Thomas Gonçalves's PhD thesis defended in September 2017 at TERATEC (Bruyères le Châtel). The main contributions of this work were also presented in the ParCo conference [14] .

6.6. BOAST: A Meta-Programming Framework to Produce Portable and Efficient Computing Kernels for HPC Applications

Participants: Brice Videau, Kevin Pouget, Luigi Genovese [CEA Inac, Grenoble], Thierry Deutsch [CEA Inac, Grenoble], Dimitri Komatitsch [LMA, CNRS, Marseille], Frédéric Desprez, Jean-Francois Mehaut.

Application portability is an important issue that should be solved efficiently, especially given the large number of different processors now available for today's supercomputers. The work needed to get performance portability is a tedious task, even for experienced programmers. The availability of semi-automatic tools is therefore mandatory for the development of large simulation applications. Computing kernels' identification and optimization has to be carefully performed as they usually consume most of the computing resources.

BOAST is a framework (DSL and run-time) that aims at describing kernels in a high-level language and it allows the comparison of the performance of different versions of the code in a simple and seamless way. We described its application to three use cases from the Mont-Blanc project. Results are encouraging as BOAST proved to be a powerful and flexible tool that allowed gains in performance compared to hand-tuned codes. Performance portability of those codes is also improved.

This work was mainly developed by Brice Videau in the context of the Mont-Blanc FP7 European projects 8.3.1.1 . It lead to a publication in the International Journal of High Performance Computing Applications (IJHPCA) [11] . A paper will also be published in 2018 describing the BOAST usage for the Gysela Application (see http://gyseladoc.gforge.inria.fr/).

6.7. Auto-tuning at Run-time with Multiple Implementations of OpenMP Tasks

Participants: Luis Felipe Garlet Millani, Lucas Mello Schnorr [UFRGS, Brazil], Jean-François Mehaut.

OpenMP established itself as the de facto standard for parallel programming in shared memory environments. It received many additions over the years enabling OpenMP to be used with heterogeneous systems. We propose an extension to the task pragma of OpenMP allowing it to provide multiple ways to compute the desired result. The run-time can thus be provided with implementations with different trade-offs.

With the use of the BOAST 5.7 auto-tuning framework, these implementations can be generated automatically before the execution. But within this framework, the auto-tuned kernel is selected in an environment different from that of an actual execution of the application. As a consequence, it may be the case that no interactions occur between different tasks during the auto-tuning, while, in the actual execution, tasks do affect each other due to shared resources like cache or memory bandwidth: Kernel selection done in isolation during the auto-tuning process is probably not the best choice for the embedded execution as part of the full application.

We propose dealing with this limitation by having the auto-tuning phase select not a single but a set of implementations, to be later further selected during execution. Our approach also permits the tuning of different parameters (such as memory accesses and number of operations), and allows to use whichever implementation is more adequate for the thread based on monitored load.

Our extension is implemented within the LLVM framework and Clang compiler front-end. Furthermore we extend the LLVM OpenMP Run-time to be aware of the multiple task implementations. We verify the efficacy of our proposal with the Ondes3D seismic wave simulator and a sparse matrix multiplication application.

6.8. Improving Characterization of NUMA Architectures through Applications' Kernels

Participants: Philippe Virouleau, Francois Broquedis, Thierry Gautier, Julien Langou [UCD, USA], Fabrice Rastello.

Programmers need tools to be able to study their applications. When targeting NUMA architectures, many existing tools allow to observe and identify the critical parts of the application. However there is a need for tools that enable programmers to clearly understand how critical parts of their applications behave, and how they could be improved on a given architecture.

In the context of data-flow applications each part - *task* - of the application is clearly identified in the data-flow graph. All manipulated data are also clearly available as, within such framework, they constitute what links tasks with one another.

On NUMA architectures, a task's execution time depends, among others, on both the core which executes the task and the NUMA node on which has been allocated its data. Assume one can characterize a task behavior (with regard to its execution context) as follow: run it in isolation from the overall application, and change various of its properties (such as the size of input or the placement of data). Then the scheduler of a run-time system can use this characteristic to improve the overall performance: It would have full information about what is running on the machine (e.g.: on the same NUMA node as the idle thread), and could sort the tasks ready for execution according to how good their behavior would be on the idle thread, given the current state.

We designed a tool which goal is to help the user execute a given *scenario* on the architecture. This scenario describes:

- What are the data and where to allocate them on the architecture
- What are the tasks to execute, where to execute them on the architecture, and with which data
- What are the characteristics to observe for each task (execution time, performance counters, ...)

The tool guarantees that the scenario will be executed correctly on the architecture, letting him focus on understanding on his application rather than taking care of the low level implementation details.

We applied this approach to a dense linear algebra algorithm: the Cholesky factorization. It has enabled us to profile the four kernels of the application by running them in various configurations of data placements, sizes, and concurrent workload. We believe we've tested enough configurations to reliably find the best and worst cases for all the kernels Assuming the behavior of the kernel stays the same within the application, we've been able to estimate upper-bound and lower-bound execution time for the overall application given those best and worst cases.

6.9. Workload-aware Loop Scheduling of Irregular Loops

Participants: Pedro Henrique de Mello Morado Penna, Marcio Castro [UFSC, Universidade Federal de Santa Catarina, Brazil], Henrique Cota de Freitas [Pontifical Catholic University of Minas Gerais, Brazil], Francois Broquedis, Jean-Francois Mehaut.

The input workload of an irregular application must be evenly distributed among its threads to enable cutting-edge performance. To address this need in OpenMP, several loop scheduling strategies were proposed. While having this ever-increasing number of strategies at disposal is helpful, it has become a non-trivial task to select the best one for a particular application. Nevertheless, this challenge becomes easier to be tackled when existing scheduling strategies are extensively evaluated. Therefore, we present a performance and scalability evaluation of the recently proposed loop scheduling strategy named Smart Round-Robin (SRR). To deliver a comprehensive analysis, we coupled a synthetic kernel benchmarking technique with several rigorous statistical tools, and considered OpenMP's Static and Dynamic loop schedulers as our baselines. Our results unveiled that SRR performs better on irregular applications with symmetric workloads and coarse-grained parallelization, achieving up to 1.9x and 1.5x speedup over OpenMP's Static and Dynamic schedulers, respectively.

This work laid the foundations of a collaboration between CORSE, the UFSC, and PUC Minas, that led to the publication of two conference papers (ICCS'17 [20], WSCAD'17 [27]) and one international journal paper (CCPE'17 [5]). The WSCAD'17 paper has also been selected to be extended for a special issue of the CCPE journal. This extension will be based on recent works with the MHM (Multiscale Hybrid-Mixed Methods) simulator developed at LNCC for the H2020 HPC4e project 8.3.1.3 between Europe and Brazil.

6.10. Error-Rate Prediction and Radiation Experiments on a 28nm Many-Core Processor

Participants: Vanessa Vargas [TIMA Labs Grenoble & ESPE, Ecuador], Pablo Ramos [TIMA Labs Grenoble & ESPE, Ecuador], Vincent Ray [kalray, Montbonnot Saint-Martin], Camille Jalier [kalray, Montbonnot Saint-Martin], Renaud Stevens [kalray, Montbonnot Saint-Martin], Benoît Dupont de Dinechin [kalray, Montbonnot Saint-Martin], Maud Baylac [LSPC Labs, Université Grenoble Alpes, CNRS/IN2P3], Francesca Villa [LSPC Labs, Université Grenoble Alpes, CNRS/IN2P3], Solenne Rey [LSPC Labs, Université Grenoble Alpes, CNRS/IN2P3], Nacer-Eddine Zergainoh [TIMA Labs, Université Grenoble Alpes & CNRS], Jean-Francois Mehaut, Raoul Velazco [TIMA Labs, Université Grenoble Alpes & CNRS].

This work analyses the 14 MeV neutron sensitivity of the MPPA-256 many-core processor. Analysis results suggest that ECC and interleaving implemented in the shared memories of clusters are very effective to mitigate SEUs as all detected events of this type were corrected.

The evaluation of the device dynamic response shows that by enabling the cache memories, it is possible to gain in performance of the application without compromising reliability, as all the detected errors produced in data and instruction cache memories were corrected by the parity protection. The non-correctable errors that occurred in the different dynamic tests were produced by bit-flips in general purpose registers since registers do not implement any protection mechanism.

Results show that the predicted application error-rate is reasonably close to the measured one. Consequently, despite the complexity of the many-core processor, this work supports the relevance of the use of the CEU approach to predict the error-rate of applications implemented in such devices.

This work is part of the Vanessa Vargas's PhD thesis defended in April 2017. This PhD thesis was advised by Raoul Velazco (TIMA) and Jean-François Méhaut. Four of the authors worked for the Kalray company. The experiments under radiation were performed on the GENEPI2 platform of the LSPC laboratory. This work was also published in the IEEE Transactions on Nuclear Sciences [10]. It was partially funded by the STIC-Amsud EnergySFE project 8.4.2 .

6.11. CAP Bench: A Benchmark Suite for Low-Power Many-Core Processors

Participants: Matheus Souza [PUC Minas], Pedro Henrique de Mello Morado Penna, Matheus Queiroz [PUC Minas], Alyson Pereira [UFSC], Luis Góes [PUC Minas], Henrique Cota de Freitas [PUC Minas], Márcio Castro [UFSC], Philippe Navaux [UFRGS], Jean-Francois Mehaut.

CAP Bench is an open source benchmark suite that includes parallel applications suitable to evaluate emerging low-power many-core processors such as MPPA-256. The benchmark contains a diverse set of applications that evaluated key aspects of MPPA-256, namely the use of its compute clusters, I/O subsystem, NoC and energy consumption. We expose development difficulties and potential bottlenecks that can stem from the shift in development paradigm when programming for low-power many-core architectures. The results showed us that different applications can have different performance bottlenecks, which is why a solid knowledge about the low-power many-core architecture is necessary for the development of efficient programs.

Our analysis shows that CAP Bench is prepared for the analysis of low-power many-core processors such as the MPPA-256, being scalable and concerned with new trends on this type of architectures. To achieve good performance and scalability, we developed applications considering aspects such as parallel patterns, load balance and architecture limitations. This allowed us to evaluate several aspects of the MPPA-256.

Our benchmark explores the hybrid programming model, which is a trend in low-power many- core processors, following parallel patterns. This enables us to verify that, in the case of MPPA-256, communication time may surpass computation time, which would ideally never occur. This behavior was highlighted by the LU application available in CAP Bench, which may indicate that the NoC should be improved to achieve better performance on NoC-bound applications. In this manner, CAP Bench comes up with the proposal to identify such bottlenecks, revealing potential improvements that might be done in future many-core architectures.

Application development challenges are still out there, and have to be solved to enable the evaluation of next generation many-core processors. As future work, we intend to incorporate other applications to the benchmark, to make it more diverse and allow for a better characterization of the architecture and its aspects. We also intend to extend the benchmark use to other many-core architectures, to achieve a broader understanding of them and the differences between many- core processors

This work was developed in the context of the EnergySFE STIC Amsud project 8.4.2 . A description of CAP Bench has been published in the CCPE (Concurrency Computation: Practice and Experience) international journal [9] . CAP Bench will be used and extended during Pedro Henrique Penna's doctoral thesis.

6.12. Social Network Analysis on Multi-Core Architectures

Participants: Thomas Messi Nguele, Maurice Tchuente [Univ Yaoundé 1, LIRIMA], Jean-Francois Mehaut.

One of social graph properties is the community structure, that is, subsets where nodes belonging to the same subset have a higher link density between themselves and a low link density with nodes belonging to external subsets. Furthermore, most social network mining algorithms comprise a local exploration of the underlying graph, which consists in referencing nodes in the neighborhood of a particular node.

The main contribution of this work is to use the community structure during the storage of large graphs that arise in social network mining. The goal is to reduce cache misses and consequently, execution time. After formalizing the problem of social network ordering as a problem of optimal linear arrangement which is known as NP-Complete, we propose NumBaCo, a heuristic based on the community structure. We present, for Katz score and Pagerank, simulations that compare classic data structures Bloc and Yale to their corresponding versions that use NumBaCo. Results on a 32-cores NUMA machine using real datasets (amazon, dblp and web-google) show that NumBaCo allows to reduce from 62% to 80% of cache misses and from 15% to 50% of execution time.

This work was initiated inside the LIRIMA Inria International Laboratory with the University of Yaoundé and Maurice Tchuenté. Those results are part of Thomas Messi Nguélé' PhD which is prepared with a Cotutelle agreement. Those results have been presented at the ParCO international conference [19] and published in the ARIMA (Revue Africaine de Recherche en Informatique et Mathématiques Appliquées) journal [3].

6.13. Run-Time Enforcement Using Büchi Games

Participants: Matthieu Renard [LaBRI], Antoine Rollet [LaBRI], Yliès Falcone.

In this work, we leverage Büchi games for the run-time enforcement of regular properties with uncontrollable events. Run-Time enforcement consists in modifying the execution of a running system to have it satisfy a given regular property, modeled by an automaton. We revisit run-time enforcement with uncontrollable events and propose a framework where we model the run-time enforcement problem as a Büchi game and synthesize sound, compliant, and optimal enforcement mechanisms as strategies. We present algorithms and a tool implementing enforcement mechanisms.We reduce the complexity of the computations performed by enforcement mechanisms at run-time by pre-computing the decisions of enforcement mechanisms ahead of time.

This work has been presented at the 24th ACM/SIGSOFT International SPIN Symposium on Model Checking of Software, SPIN 2017 [23].

6.14. GREP: Games for the Run-Time Enforcement of Properties

Participants: Matthieu Renard [LaBRI], Antoine Rollet [LaBRI], Yliès Falcone.

In this work, we developed GREP, a tool for the run-time enforcement of (timed) properties. GREP takes an execution sequence as input (stdin), and modifies it (stdout) as necessary to enforce the desired property, when possible. GREP can enforce any regular timed property described by a deterministic and complete Timed Automaton. The main novelties of GREP are twofold: It uses game theory to improve the synthesis of enforcement mechanisms, and it accounts for uncontrollable events, i.e. events that cannot be controlled by the enforcement mechanisms and thus have to be released immediately. The usability of GREP has been validated with a performance evaluation.

The associated work has been presented at the IFIP International Conference on Testing Software and Systems, ICTSS 2017 [22]

6.15. Verifying Policy Enforcers

Participants: Oliviero Riganelli [University of Milano Bicocca], Daniela Micucci [University of Milano Bicocca], Leonardo Mariani [University of Milano Bicocca], Yliès Falcone.

Policy enforcers are sophisticated run-time components that can prevent failures by enforcing the correct behavior of the software. While a single enforcer can be easily designed focusing only on the behavior of the application that must be monitored, the effect of multiple enforcers that enforce different policies might be hard to predict. So far, mechanisms to resolve interferences between enforcers have been based on priority mechanisms and heuristics. Although these methods provide a mechanism to take decisions when multiple enforcers try to affect the execution at a same time, they do not guarantee the lack of interference on the global behavior of the system. In this work we propose a verification strategy that can be exploited to discover interferences between sets of enforcers and thus safely identify a priori the enforcers that can co-exist at run-time. In our evaluation, we experimented our verification method with several policy enforcers for Android and discovered some incompatibilities.

This work has been presented at the 17-th International Conference on Run-Time Verification, RV 2017 [24].

6.16. Monitoring Decentralized Specifications

Participants: Antoine El-Hokayem, Yliès Falcone.

In this work, we define two complementary approaches to monitor decentralized systems. The first approach relies on those with a centralized specification, i.e, when the specification is written for the behavior of the entire system. To do so, our approach introduces a data-structure that i) keeps track of the execution of an automaton, ii) has predictable parameters and size, and iii) guarantees strong eventual consistency. The second approach defines decentralized specifications wherein multiple specifications are provided for separate parts of the system. We study decentralized monitorability, and present a general algorithm for monitoring decentralized specifications. We map three existing algorithms to our approaches and provide a framework for analyzing their behavior.

The associate tool THEMIS 5.1 is a framework for designing such decentralized algorithms, and simulating their behavior. This work has been presented at the 26th ACM/SIGSOFT International Symposium on Software Testing and Analysis, ISSTA 2017 [12].

6.17. Interactive Run-Time Verification - When Interactive Debugging Meets Run-Time Verification

Participants: Raphael Jakse, Yliès Falcone, Kevin Pouget, Jean-Francois Mehaut.

Run-Time Verification consists in studying a system at run-time, looking for input and output events to discover, check or enforce behavioral properties. Interactive debugging consists in studying a system at run-time in order to discover and understand its bugs and fix them, inspecting interactively its internal state. Interactive Run-Time Verification (i-RV) combines run-time verification and interactive debugging. We define an efficient and convenient way to check behavioral properties automatically on a program using a debugger. We aim at helping bug discovery and understanding by guiding classical interactive debugging techniques using run-time verification.

This work has been presented at the IEEE 28th International Symposium on Software Reliability Engineering, ISSRE 2017 [15]. It is also a part of the Nano2017 DEMA project 7.2 with STMicroelectronics.

6.18. Predictive Run-Time Verification of Timed Properties

Participants: Srinivas Pinisetty [Aalto University], Thierry Jéron [Inria Rennes], Stravos Tripakis [Aalto University], Yliès Falcone, Hervé Marchand [Inria Rennes], Viorel Preoteasa [Aalto University].

Run-Time verification (RV) techniques are used to continuously check whether the (un-trustworthy) output of a black-box system satisfies or violates a desired property. When we consider run-time verification of timed properties, physical time elapsing between actions influences the satisfiability of the property. This work introduces predictive run-time verification of timed properties where the system is not entirely a black-box but something about its behavior is known a priori. A priori knowledge about the behavior of the system allows the verification monitor to foresee the satisfaction (or violation) of the monitored property. In addition to providing a conclusive verdict earlier, the verification monitor also provides additional information such as the minimum (maximum) time when the property can be violated (satisfied) in the future. The feasibility of the proposed approach is demonstrated by a prototype implementation, which is able to synthesize predictive run-time verification monitors from timed automata.

This work has been published in the Journal of Systems and Software 2017 [6].

6.19. Concurrency-Preserving and Sound Monitoring of Multi-Threaded Component-based Systems: Theory, Algorithms, Implementation, and Evaluation

Participants: Hosein Nazarpour [Verimag], Yliès Falcone, Saddek Bensalem [Verimag], Marius Bozga [Verimag].

This work addresses the monitoring of logic-independent linear-time user-provided properties in multi-threaded component-based systems. We consider intrinsically independent components that can be executed concurrently with a centralized coordination for multiparty interactions. In this context, the problem that arises is that a global state of the system is not available to the monitor. A naive solution to this problem would be to plug in a monitor which would force the system to synchronize in order to obtain the sequence of global states at run-time. Such a solution would defeat the whole purpose of having concurrent components. Instead, we reconstruct on-the-fly the global states by accumulating the partial states traversed by the system at run-time. We define transformations of components that preserve their semantics and concurrency and, at the same time, allow to monitor global-state properties. Moreover, we present RVMT-BIP, a prototype tool implementing the transformations for monitoring multi-threaded systems described in the Behavior, Interaction, Priority (BIP) framework, an expressive framework for the formal construction of heterogeneous systems. Our experiments on several multi-threaded BIP systems show that RVMT-BIP induces a cheap run-time overhead.

This work has been published in the journal Formal Aspects of Computing 2017 [4].

6.20. Formal Analysis and Offline Monitoring of Electronic Exams

Participants: Ali Kassem [Inria Grenoble], Yliès Falcone, Pascal Lafourcade [University of Auvergne].

More and more universities are moving toward electronic exams (in short e-exams). This migration exposes exams to additional threats, which may come from the use of the information and communication technology. In this work, we identify and define several security properties for e-exam systems. Then, we show how to use these properties in two complementary approaches: model-checking and monitoring. We illustrate the validity of our definitions by analyzing a real e-exam used at the pharmacy faculty of University Grenoble Alpes (UGA) to assess students. On the one hand, we instantiate our properties as queries for ProVerif, an automatic verifier of cryptographic protocols, and we use it to check our modeling of UGA exam specifications. ProVerif found some attacks. On the other hand, we express our properties as Quantified Event Automata (QEAs), and we synthesize them into monitors using MarQ, a Java tool designed to implement QEAs. Then, we use these monitors to verify real exam executions conducted by UGA. Our monitors found fraudulent students and discrepancies between the specifications of UGA exam and its implementation.

This work has been published in the journal of Formal Methods in System Design 2017 [2].

6.21. Teaching Algorithms using Problem and Challenge Based Learning

Participant: Florent Bouchez - Tichadou.

Teaching algorithms is always a challenge at any level of the CS curriculum, as it is often viewed as a theoretical field. While many exercises revolve around classical examples that illustrate interesting algorithmic points, they are often disconnected from reality, which is a major drawback for students trying to learn. During the last four years, we have been trying to reconnect the teaching of algorithms with their applicability in the real world to M1 and L2 students, by giving them actual problems that could arise in their life of future software engineers, challenging enough to force them to use particular algorithmic techniques or data structures—e.g., linked lists, binary trees, dynamic programming or approximation algorithms.

By assigning students in groups of 5 to 6 members, we wanted to create an environment where they function as a team trying to work together to solve a problem. This allowed them to help each other in their respective comprehension, and made them more autonomous in their learning. The effective materials was provided as online pdf files so they had to read and learn from them by themselves, while the class sessions with a tutor (teacher) where used for the problem-solving part, with guidance from the tutor (who is there to make sure the learning takes place).

After four years of experimentation with M1 students, we found that the student's grades were stable, in particular there was no decrease in exams' performances compared to the classical course that was taught in the previous years. However, the students progressed in trans-disciplinary skills such a communication and the writing of essays. More importantly, students show a strong adhesion to the teaching method, 50% of them rating it as "excellent" (6) and 25% as "good" (resp. 6 and 5 on a scale from 1 (terrible) to 6 (excellent)). No student rated the course below average.

This work is still ongoing, and our plan now is to use our knowledge of the internals of compilers and run-time systems to: First, extract real-life algorithmic problems that have concrete applications; Second, create a tool that exposes the working mechanics of a running program, hence helping students to better understand how algorithms work.

PACAP Project-Team

6. New Results

6.1. Compiler, vectorization, interpretation

Participants: Erven Rohou, André Seznec, Sylvain Collange, Rabab Bouziane, Arif Ali Ana-Pparakkal, Stefano Cherubin, Byron Hawkins, Arif Ali Ana-Pparakkal, Imane Lasri, Kévin Le Bon.

6.1.1. Improving sequential performance through memoization

Participants: Erven Rohou, Imane Lasri, André Seznec.

Many applications perform repetitive computations, even when properly programmed and optimized. Performance can be improved by caching results of pure functions, and retrieving them instead of recomputing a result (a technique called memoization).

We previously proposed [23] a simple technique for enabling software memoization of any dynamically linked pure function and we illustrate our framework using a set of computationally expensive pure functions – the transcendental functions.

A restriction of the proposed framework was that memoization was restricted only to dynamically linked functions and the functions must be determined beforehand. We extended this work, and we propose function memoization using a compile-time technique thus extending the scope of memoization to user defined functions as well as making it transparently applicable to any dynamically linked functions. Our compile-time technique allows static linking of memoization code and this increases the benefit due to memoization by leveraging the inlining capability for the memoization wrapper. Our compile-time analysis can also handle functions with pointer parameters , and we handle constants more efficiently. Instruction set support can also be considered, and we propose associated hardware leading to additional performance gain.

This work was presented at the Compiler Construction Conference 2017 [50]. It is also described in the PhD thesis of Arjun Suresh [24].

6.1.2. Optimization in the Presence of NVRAM

Participants: Erven Rohou, Rabab Bouziane.

Beyond the fact of generating machine code, compilers play a critical role in delivering high performance, and more recently high energy efficiency. For decades, the memory technology of target systems has consisted in SRAM at cache level, and DRAM for main memory. Emerging non-volatile memories (NVMs) open up new opportunities, along with new design challenges. In particular, the asymmetric cost of read/write accesses calls for adjusting existing techniques in order to efficiently exploit NVMs. In addition, this technology makes it possible to design memories with cheaper accesses at the cost of lower data retention times. These features can be exploited at compile time to derive better data mappings according to the application and data retention characteristics. We reviewed a number of compile-time analysis and optimization techniques, and how they could apply to systems in presence of NVMs [37]. In particular, we consider the case of the reduction of the number of writes, and the analysis of variables lifetime for memory bank assignment of program variables.

Concerning the reduction of writes, we propose a fast evaluation of NVM integration at cache level, together with a compile-time approach for mitigating the penalty incurred by the high write latency of STT-RAM. We implement a code optimization in LLVM for reducing so-called *silent stores*, i.e., store instruction instances that write to memory values that were already present there. This makes our optimization portable over any architecture supporting LLVM. Then, we assess the possible benefit of such an optimization on the Rodinia benchmark suite through an analytic approach based on parameters extracted from the literature devoted to NVMs. This makes it possible to rapidly analyze the impact of NVMs on memory energy consumption. Reported results show up to 42 % energy gain when considering STT-RAM caches. This work is accepted for publication at RAPIDO'18 [38].

This research is done in collaboration with Abdoulaye Gamatié at LIRMM (Montpellier) within the context the the ANR project CONTINUUM.

6.1.3. Dynamic Binary Optimization

Participants: Erven Rohou, Arif Ali Ana-Pparakkal, Kévin Le Bon, Byron Hawkins.

6.1.3.1. Dynamic Function Specialization

Participants: Erven Rohou, Arif Ali Ana-Pparakkal, Kévin Le Bon.

Compilers can do better optimization with the knowledge of run-time behavior of the program. *Function specialization* is a compilation technique that consists in optimizing the body of a function for specific values of an argument. Different versions of a function are created to deal with the most frequent values of the arguments, as well as the default case. Compilers can do a better optimization with the knowledge of run-time behaviour of the program. Static compilers, however, can hardly predict the exact value/behaviour of arguments, and even profiling collected during previous runs is never guaranteed to capture future behaviour. We propose a dynamic function specialization technique, that captures the actual values of arguments during execution of the program and, when profitable, creates specialized versions and include them at runtime. Our approach relies on dynamic binary rewriting. We present [36] the principles and implementation details of our technique, analyze sources of overhead, and present our results.

This research is done within the context of the Nano 2017 PSAIC collaborative project.

6.1.3.2. Runtime Vectorization of Binary Programs

Participant: Erven Rohou.

In many cases, applications are not optimized for the hardware on which they run. Several reasons contribute to this unsatisfying situation, such as legacy code, commercial code distributed in binary form, or deployment on compute farms. In fact, backward compatibility of ISA guarantees only the functionality, not the best exploitation of the hardware. In this work, we focus on maximizing the CPU efficiency for the SIMD extensions.

We previously proposed [3] a binary-to-binary optimization framework where loops vectorized for an older version of the processor SIMD extension are automatically converted to a newer one. It is a lightweight mechanism that does not include a vectorizer, but instead leverages what a static vectorizer previously did. We showed that many loops compiled for x86 SSE can be dynamically converted to the more recent and more powerful AVX; as well as, how correctness is maintained with regards to challenges such as data dependencies and reductions. We obtained speedups in line with those of a native compiler targeting AVX.

We now focus on runtime vectorization of loops in binary codes that were not originally vectorized [29]. For this purpose, we use open source frameworks that we have tuned and integrated to

1. dynamically lift the x86 binary into the Intermediate Representation form of the LLVM compiler,
2. abstract hot loops in the polyhedral model,
3. use the power of this mathematical framework to vectorize them,
4. and finally compile them back into executable form using the LLVM Just-In-Time compiler.

In most cases, the obtained speedups are close to the number of elements that can be simultaneously processed by the SIMD unit. The re-vectorizer and auto-vectorizer are implemented inside a dynamic optimization platform; it is completely transparent to the user, does not require any rewriting of the binaries, and operates during program execution.

This work is done in collaboration with Philippe Clauss (Inria CAMUS), it is part of the PhD work of Nabil Hallou [26].

6.1.4. Hardware/Software JIT Compiler

Participant: Erven Rohou.

Dynamic Binary Translation (DBT) is often used in hardware/software co-design to take advantage of an architecture model while using binaries from another one. The co-development of the DBT engine and of the execution architecture leads to architecture with special support to these mechanisms. We proposed [46] a hardware accelerated dynamic binary translation where the first steps of the DBT process are fully accelerated in hardware. Results showed that using our hardware accelerators leads to a speed-up of 8× and a cost in energy 18× lower, compared with an equivalent software approach.

Single ISA-Heterogeneous multi-cores such as the ARM big.LITTLE have proven to be an attractive solution to explore different energy/performance trade-offs. Such architectures combine Out of Order cores with smaller in-order ones to offer different power/energy profiles. They however do not really exploit the characteristics of workloads (compute-intensive vs. control dominated). In our recent work, we propose to enrich these architectures with runtime configurable VLIW cores, which are very efficient at compute-intensive kernels. To preserve the single ISA programming model, we resort to Dynamic Binary Translation, and use this technique to enable dynamic code specialization for Runtime Reconfigurable VLIWs cores. Our proposed DBT framework targets the RISC-V ISA, for which both OoO and in-order implementations exist. Our experimental results show that our approach can lead to best-case performance and energy efficiency when compared against static VLIW configurations.

This work has been accepted for publication at DATE 2018 [53].

This research is done in collaboration with Steven Derrien and Simon Rokicki from the CAIRN team.

6.1.5. Customized Precision Computing

Participants: Erven Rohou, Stefano Cherubin, Imane Lasri.

Error-tolerating applications are increasingly common in the emerging field of real-time HPC. Proposals have been made at the hardware level to take advantage of inherent perceptual limitations, redundant data, or reduced precision input, as well as to reduce system costs or improve power efficiency. At the same time, works on floating-point to fixed-point conversion tools allow us to trade-off the algorithm exactness for a more efficient implementation. In this work [39], we aim at leveraging existing, HPC-oriented hardware architectures, while including in the precision tuning an adaptive selection of floating-and fixed-point arithmetic. Our proposed solution takes advantage of the application domain knowledge of the programmers by involving them in the first step of the interaction chain. We rely on annotations written by the programmer on the input file to know which variables of a computational kernel should be converted to fixed-point. The second stage replaces the floating-point variables in the kernel with fixed-point equivalents. It also adds to the original source code the utility functions to perform data type conversions from floating-point to fixed-point, and vice versa. The output of the second stage is a new version of the kernel source code which exploits fixed-point computation instead of floating-point computation. As opposed to typical custom-width hardware designs, we only rely on the standard 16-bit, 32-bit and 64-bit types. We also explore the impact of the fixed-point representation on auto-vectorization. We discuss the effect of our solution in terms of time-to-solutions, error and energy-to-solution.

This is done within the context of the ANTAREX project in collaboration with Stefano Cherubin, and Giovanni Agosta from Politecnico di Milano, and Olivier Sentieys from the CAIRN team.

6.1.6. SPMD Function Call Re-Vectorization

Participant: Sylvain Collange.

SPMD programming languages for SIMD hardware such as C for CUDA, OpenCL or ISPC have contributed to increase the programmability of SIMD accelerators and graphics processing units. However, SPMD languages still lack the flexibility offered by low-level SIMD programming on explicit vectors. To close this expressiveness gap while preserving the SPMD abstraction, we introduce the notion of Function Call Re-Vectorization (CREV). CREV allows changing the dimension of vectorization during the execution of an SPMD kernel, and exposes it as a nested parallel kernel call. CREV affords a programmability close to dynamic parallelism, a feature that allows the invocation of kernels from inside kernels, but at much lower cost. We defined a formal semantics of CREV, and implemented it on the ISPC compiler. To validate our

idea, we have used CREV to implement some classic algorithms, including string matching, depth first search and Bellman-Ford, with minimum effort. These algorithms, once compiled by ISPC to Intel-based vector instructions, are as fast as state-of-the-art implementations, yet much simpler. As an example, our straightforward implementation of string matching beats the Knuth-Morris-Pratt algorithm by 12%. This work was presented at the ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming (PPoPP) 2017 [45].

This work was done in collaboration with Rubens Emilio and Fernando Pereira at UFMG, as part of the Inria PROSPIEL Associate Team.

6.1.7. Qubit allocation for quantum circuit compilers

Participant: Sylvain Collange.

Quantum computing hardware is becoming a reality. For instance, IBM Research makes a quantum processor available in the cloud to the general public. The possibility of programming an actual quantum device has elicited much enthusiasm. Yet, quantum programming still lacks the compiler support that modern programming languages enjoy today. To use universal quantum computers like IBM's, programmers must design low-level circuits. In particular, they must map logical qubits into physical qubits that need to obey connectivity constraints. This task resembles the early days of programming, in which software was built in machine languages. We have formally introduced the qubit allocation problem and provided an exact solution to it. This optimal algorithm deals with the simple quantum machinery available today; however, it cannot scale up to the more complex architectures scheduled to appear. Thus, we also provide a heuristic solution to qubit allocation, which is faster than the current solutions already implemented to deal with this problem.

This paper is accepted for publication at the Code Generation and Optimization (CGO) conference [49].

This work was done in collaboration with Vinícius Fernandes dos Santos, Fernando Pereira and Marcos Yukio Siraichi at UFMG, Brazil.

6.2. Processor Architecture

Participants: Pierre Michaud, Sylvain Collange, Erven Rohou, André Seznec, Biswabandan Panda, Fernando Endo, Kleovoulos Kalaitzidis, Daniel Rodrigues Carvalho, Anita Tino.

Processor, cache, locality, memory hierarchy, branch prediction, multicore, power, temperature

6.2.1. Microarchitecture

6.2.1.1. Bayesian TAGE predictors

Participant: Pierre Michaud.

The TAGE conditional branch predictor, introduced by André Seznec and Pierre Michaud in 2006, is the most storage-efficient branch predictor known today [19]. André Seznec has won the last four branch prediction championships, each time with a TAGE-based predictor. However, since 2006, the improvements in prediction accuracy have been relatively modest and were mostly obtained at the cost of increased hardware complexity. In particular, André Seznec added a Statistical Corrector to TAGE to correct some of its deficiencies [21]. This may be an indication that our understanding of TAGE is not complete and that further accuracy gains are waiting to be discovered. The problem tackled by the statistical corrector is that of cold counters: a TAGE-like predictor constantly allocate new entries, erasing the branch history information stored in the up-down counters of the overwritten entries. TAGE mitigates this problem by using the confidence level of the up-down counter and a meta-predictor. However, fundamentally, the information on the degree of coldness of the up-down counter is not available in TAGE. Therefore we propose to replace the up-down counter with a dual-counter counting separately taken and not-taken occurrences. Replacing the up-down counter with a dual-counter requires to redefine prediction confidence estimation. We found that a Bayesian formula, namely Laplace's rule of succession, provides effective confidence estimation. We also discovered a method, based on the dual-counter, for reducing the number of allocations. By combining these new findings, we devised a new TAGE-like predictor called BATAGE, more accurate than TAGE, making external statistical correction superfluous. As of December 2017, this work is in the process of being submitted to a journal.

6.2.1.2. Interactions Between Value Prediction and Compiler Optimizations
Participants: André Seznec, Fernando Endo.

Increasing instruction-level parallelism is regaining attractiveness within the microprocessor industry. The EOLE microarchitecture [13] and D-VTAGE value predictor [14] were recently introduced to solve practical issues of value prediction (VP). In particular, they remove the most significant difficulties that forbade an effective VP hardware. In [28], we present a detailed evaluation of the potential of VP in the context of EOLE/D-VTAGE and different compiler options. Our study shows that if no single general rule always applies – more optimization might sometimes leads to more performance – unoptimized codes often gets a large benefit from the prediction of redundant loads.

6.2.1.3. Prefetch Management on Multicore Systems
Participants: André Seznec, Biswabandan Panda.

In multi-core systems, an application's prefetcher can interfere with the memory requests of other applications using the shared resources, such as last level cache and memory bandwidth. Towards this end, we propose a solution to manage prefetching in multi-core systems [32]. In particular, we make two fundamental observations: First, a strong positive correlation exists between the accuracy of a prefetcher and the amount of prefetch requests it generates relative to an application's total (demand and prefetch) requests. Second, a strong positive correlation exists between the ratio of total prefetch to demand requests and the ratio of average last level cache miss service times of demand to prefetch requests. In [32], we propose Band-pass prefetching a simple and low-overhead mechanism to effectively manage prefetchers in multi-core systems that builds on those two observations. Our solution consists of local and global prefetcher aggressiveness control components, which altogether, control the flow of prefetch requests between a range of prefetch to demand requests ratios.

6.2.1.4. Managing Shared Last Level Caches in Large Multicores
Participant: André Seznec.

Multi-core processors employ shared Last Level Caches (LLC). This trend continuewith large multi-core processors (16 cores and beyond) as well. At the same time, the associativity of LLC tends to remain in the order of sixteen. Consequently, with large multicore processors, the number of applications or threads that share the LLC becomes larger than the associativity of the cache itself. LLC management policies have been extensively studied for small scale multi-cores (4 to 8 cores) and associativity degree in the 16 range. However, the impact of LLC management on large multi-cores is essentially unknown, in particular when the associativity degree is smaller than the number of applications. In [33], we introduce Adaptive Discrete and deprioritized Application PrioriTization (ADAPT), an LLC management policy addressing the large multi-cores where the LLC associativity degree is smaller than the number of applications. ADAPT builds on the use of the Footprint-number metric. We propose a monitoring mechanism that dynamically samples cache sets to estimate the Footprint-number of applications and classifies them into discrete (distinct and more than two) priority buckets. The cache replacement policy leverages this classification and assigns priorities to cache lines of applications during cache replacement operations. We further find that deprioritizing certain applications during cache replacement is beneficial to the overall performance.

6.2.1.5. Augmenting superscalar architecture for efficient many-thread parallel execution
Participants: Sylvain Collange, André Seznec.

Threads of Single-Program Multiple-Data (SPMD) applications often exhibit very similar control flows, i.e. they execute the same instructions on different data. We propose the Dynamic Inter-Thread Vectorization Architecture (DITVA) to leverage this implicit data-level parallelism in SPMD applications by assembling dynamic vector instructions at runtime. DITVA extends an in-order SMT processor with SIMD units with an inter-thread vectorization execution mode. In this mode, multiple scalar threads running in lockstep share a single instruction stream and their respective instruction instances are aggregated into SIMD instructions. To balance thread-and data-level parallelism, threads are statically grouped into fixed-size independently scheduled warps. DITVA leverages existing SIMD units and maintains binary compatibility with existing CPU

architectures. Our evaluation on the SPMD applications from the PARSEC and Rodinia OpenMP benchmarks shows that a 4-warp × 4-lane 4-issue DITVA architecture with a realistic bank-interleaved cache achieves 1.55× higher performance than a 4-thread 4-issue SMT architecture with AVX instructions while fetching and issuing 51 % fewer instructions, achieving an overall 24 % energy reduction.

This work has been accepted for publication in the Journal of Parallel and Distributed Computing [30]. It was done in collaboaration with Sajith Kalathingal and Bharath Swamy from Intel Bangalore (India).

6.2.1.6. Generalizing the SIMT execution model to general-purpose instruction sets

Participant: Sylvain Collange.

The *Single Instruction, Multiple Threads* (SIMT) execution model as implemented in NVIDIA Graphics Processing Units (GPUs) associates a multi-thread programming model with an SIMD execution model [57]. It combines the simplicity of scalar code from the programmer's and compiler's perspective with the efficiency of SIMD execution units at the hardware level. However, current SIMT architectures demand specific instruction sets. In particular, they need specific branch instructions to manage thread divergence and convergence. Thus, SIMT GPUs have remained incompatible with traditional general-purpose CPU instruction sets.

We designed Simty, an SIMT processor proof of concept that lifts the instruction set incompatibility between CPUs and GPUs. Simty is a massively multi-threaded processor core that dynamically assembles SIMD instructions from scalar multi-thread code. It runs the RISC-V (RV32-I) instruction set. Unlike existing SIMD or SIMT processors like GPUs, Simty takes binaries compiled for general-purpose processors without any instruction set extension or compiler changes. Simty is described in synthesizable RTL. A FPGA prototype validates its scaling up to 2048 threads per core with 32-wide SIMD units.

The Simty architecture was presented at the First Workshop on Computer Architecture Research with RISC-V (CARRV 2017) [40].

Both conventional and generalized SIMT architectures like Simty use hardware or software mechanisms to keep track of control-flow divergence and convergence among threads. A new class of such algorithms is gaining popularity in the literature in the last few years. We presented a new classification of these techniques based on their common characteristic, namely traversals of the control-flow graph based on lists of paths. We compared the implementation cost on an FPGA of path lists and per-thread program counters within the Simty processor. The sorted list enables significantly better scaling starting from 8 threads per warp.

This work was presented in French in Conférence d'informatique en Parallélisme, Architecture et Système (ComPAS) [51] and is available in English as a technical report [52].

6.2.1.7. Toward out-of-order SIMT microarchitecture

Participants: Sylvain Collange, Anita Tino.

Prior work highlights the continued importance of maintaining adequate sequential performance within throughput-oriented cores [60]. Out-of-order superscalar architectures as used in high-performance CPU cores can meet such demand for single-thread performance. However, GPU architectures based on SIMT have been limited so far to in-order execution because of a major scientific obstacle: the partial dependencies between instructions that SIMT execution induces thwart register renaming. This ongoing project is seeking to generalize out-of-order execution to SIMT architectures. In particular, we revisit register renaming techniques originally proposed for predicate conversion to support partial register updates efficiently. Out-of-order dynamic vectorization holds the promise to close the CPU-GPU design space by enabling low-latency, high-throughput design points.

6.3. WCET estimation and optimization

Participants: Isabelle Puaut, Damien Hardy, Viet Anh Nguyen, Benjamin Rouxel, Sébastien Martinez, Erven Rohou, Imen Fassi, Loïc Besnard, Stefanos Skalistis.

6.3.1. WCET estimation for many core processors

Participants: Viet Anh Nguyen, Damien Hardy, Sébastien Martinez, Isabelle Puaut, Benjamin Rouxel.

6.3.1.1. Optimization of WCETs by considering the effects of local caches

The overall goal of this research is to define WCET estimation methods for parallel applications running on many-core architectures, such as the Kalray MPPA machine.

Some approaches to reach this goal have been proposed, but they assume the mapping of parallel applications on cores already done. Unfortunately, on architectures with caches, task mapping requires a priori known WCETs for tasks, which in turn requires knowing task mapping (i.e., co-located tasks, co-running tasks) to have tight WCET bounds. Therefore, scheduling parallel applications and estimating their WCET introduce a chicken and egg situation.

We have addressed this issue by developing both optimal and heuristic techniques for solving the scheduling problem, whose objective is to minimize the WCET of a parallel application. Our proposed static partitioned non-preemptive mapping strategies address the effect of local caches to tighten the estimated WCET of the parallel application. Experimental results obtained on real and synthetic parallel applications show that co-locating tasks that reuse code and data improves the WCET by 11 % on average for the optimal method and by 9 % on average for the heuristic method [35].

This research is part of the PIA Capacités project.

6.3.1.2. Accounting for shared resource contentions to minimize WCETs

Accurate WCET analysis for multi-cores is known to be challenging, because of concurrent accesses to shared resources, such as communication through busses or Networks on Chips (NoC). Since it is impossible in general to guarantee the absence of resource conflicts during execution, current WCET techniques either produce pessimistic WCET estimates or constrain the execution to enforce the absence of conflicts, at the price of a significant hardware under-utilization. In addition, the large majority of existing works consider that the platform workload consists of independent tasks. As parallel programming is the most promising solution to improve performance, we envision that within only a few years from now, real-time workloads will evolve toward parallel programs. The WCET behavior of such programs is challenging to analyze because they consist of *dependent* tasks interacting through complex synchronization/communication mechanisms.

In a first work (thesis of Benjamin Rouxel), we proposed techniques that account for interferences to access shared ressources, in order to minimize the WCET of parallel applications. An optimal and a heuristic method are proposed to map and schedule tasks on multi-cores. These methods take the structure of applications (synchronizations/communications) into consideration to tightly identify shared resource interferences and consequently tighten WCET estimates. Our heuristic improves by 19% the overall WCET compared to a worst-case contention baseline [47], [31].

In a second study [44], we have studied the gain that could be obtained on an initially produced time-triggered non-preemptive schedule, by the introduction of slack time, in order to avoid interference between tasks. The introduction of slack time is performed using an optimal technique using Integer Linear Programming (ILP), to evaluate how much at best can be gained. Experimental results using synthetic task graphs and a Kalray-like architecture with round-robin bus arbitration show that avoiding contention reduces WCETs, albeit by a small percentage. The highest reductions are observed on applications with the highest memory demand, and when the application is scheduled on the highest number of cores.

This work is performed in cooperation with Steven Derrien from the CAIRN research group and is part of the ARGO H2020 project.

6.3.1.3. WCET-Aware Parallelization of Model-Based Applications for Multi-Cores

Parallel architectures are nowadays no longer confined to the domain of high performance computing, they are also increasingly used in embedded time-critical systems.

The ongoing ARGO H2020 project provides a programming paradigm and associated tool flow to exploit the full potential of architectures in terms of development productivity, time-to-market, exploitation of the platform computing power and guaranteed real-time performance. In [41] we give an overview of the objectives of ARGO and explore the challenges introduced by our approach.

6.3.2. WCET estimation tool and benchmarks

Participants: Damien Hardy, Isabelle Puaut, Benjamin Rouxel, Loïc Besnard.

Estimation of worst-case execution times (WCETs) is required to validate the temporal behavior of hard real time systems. Heptane is an open-source software program that estimates upper bounds of execution times on MIPS and ARM v7 architectures, offered to the WCET estimation community to experiment new WCET estimation techniques. The software architecture of Heptane was designed to be as modular and extensible as possible to facilitate the integration of new approaches. In [42], we present the current status of Heptane, give information on the analyses it implements, as well as how to use it and extend it.

We all had quite a time to find non-proprietary architecture-independent exploitable parallel benchmarks for Worst-Case Execution Time (WCET) estimation and real-time scheduling. However, there is no consensus on a parallel benchmark suite, when compared to the single-core era and the Mälardalen benchmark suite. In [48] we bridge part of this gap, by presenting a collection of benchmarks with the following good properties: (i) easily analyzable by static WCET estimation tools (written in structured C language, in particular neither goto nor dynamic memory allocation, containing flow information such as loop bounds); (ii) independent from any particular run-time system (MPI, OpenMP) or real-time operating system. Each benchmark is composed of the C source code of its tasks, and an XML description describing the structure of the application (tasks and amount of data exchanged between them when applicable). Each benchmark can be integrated in a full end-to-end empirical method validation protocol on multi-core architecture. This proposed collection of benchmarks is derived from the well known StreamIT benchmark suite and will be integrated in the TACleBench suite in a near future.

6.4. Security

Participants: Erven Rohou, Damien Hardy, Nicolas Kiss.

Physical attacks represent a very important threat in the context of embedded systems: these attacks try to recover cryptographic keys by exploiting the physical behavior of the device. They can either be passive (e.g. by monitoring the power consumption of the device) or active (e.g. by injecting errors to reveal or deduce sensitive data).

One family of countermeasures to protect against those passive attacks (also known as *side-channel* attacks) is called masking. The principle is to “hide” data with masks so that internal values used in computations can not be predicted with the behavior observed. We modified the LLVM compiler (version 3.8) to automatically insert masking coutermeasures into the code at compile-time. Our modification works at intermediate level (IR level), this way we can perform low-level transformations (e.g. memory allocation, instructions replacement) while covering most of the architectures used in the embedded world.

The main innovation of this work is the generic approach used for the transformation and thus, the ability to easily change the masking scheme without modifying the compiler internal code. We introduced a way to describe in high-level language (C/C++) the masking operations independently in what we call “primitives”. With this technique, we implemented “Boolean Masking” and we tested the efficiency on an embedded implementation of AES. After measuring the electromagnetic emissions of 20,000 executions, we performed a Correlation Power Analysis (CPA) and results have shown that the countermeasure is correctly applied. Hence, it is not possible anymore to recover the cryptographic key with this type of attack.

This work is done in the context of the SECODE CHIST-ERA project.

AOSTE2 Team

7. New Results

7.1. Uniprocessor Mixed-Criticality Real-Time Scheduling

Participants: Slim Ben-Amor, Liliana Cucu, Robert Davis, Mehdi Mezouak, Yves Sorel.

In the framework of the FUI CEOS project 9.1.1.1 we mainly investigated the PX4 autopilot free software program that was chosen by the partners to be implemented on the Pixhawk electronic board. This board will be installed in the multirotor drone that the project is intended to built. The board is based on a microcontroller which contains an ARM Cortex M4 microprocessor, timers, several sensors, accelerometer, gyroscope, magnetometer, barometer, and actuators, mainly four to eight electric motors depending on the level of redundancy.

We studied the existing source code of PX4 which consists of two main layers: the flight stack, which is an estimation and flight control system, and the middleware, which is a general robotics layer providing internal/external communications and hardware integration. This study allowed us to understand the general architecture of PX4. The flight stack is split into a set of threads communicating asynchronously through a micro object request broker messaging. In the CEOS project our team is in charge to guarantee that the drone will satisfy multiple real-time criticality levels. In order to be able to perform a real-time schedulability analysis on the PX4 autopilot, first we transformed this set of communicating threads into a task dependency graph. Second, we sought the period of each task starting from input tasks which read from sensors, to output tasks which write into actuators. The partners of the project chose to run PX4 on the NuttX OS which is open source, light-weight, efficient and very stable. It provides POSIX API and some form of real-time scheduling. Thus, we had to deeply understand the scheduler and the management of interruptions and time of NuttX. We plan to modify NuttX in order to support mixed-criticality applications using to start, online real-time scheduling, and then offline real-time scheduling.

Finally, always to perform the real-time schedulability analysis of PX4, we must estimate the worst execution time (WCET) of each task. This problem is very complex due to the multiple possible paths in a task as well as the different data it consumes. Moreover, the processor and/or the microcontroller itself may have some features like memory contentions, bus accesses, caches, pipelines, speculative branchings that increase the difficulty to determine WCETs. All these variabilities lead us to introduce probabilistic reasoning in characterizing the timing behavior (WCET, schedulability analyses) of mixed-criticality real-time applications [4].

7.2. Multiprocessor Real-Time Scheduling

Participants: Salah-Eddine Saidi, Yves Sorel.

During the third year of the PhD thesis of Salah Eddine Saidi, we focused on two aspects. First, we finalized our work on the parallelization on multi-core processors of FMI-based co-simulation of numerical models in order to accelerate its execution. Our approach, based on the transformation of FMU graphs into operation graphs which reveal more parallelism, comprises the following two steps: first acyclic orientation necessary for avoiding that some operations of a same model are executed in parallel and second multi-core offline scheduling of operations [5]. We proposed exact algorithms based on ILP (Integer Linear Programming) and heuristics for performing the acyclic orientation and the multi-core scheduling. Also, we proposed a random generator of synthetic co-simulations. Using these generated co-simulations, we compared the performances of the heuristics and the ILP-based exact algorithm for both the acyclic orientation and the scheduling in terms of execution time and quality of the obtained solution. Tests have been carried out for different sizes of co-simulation and different numbers of cores. Moreover, we compared the performance of our offline approach with an online scheduling approach based on the Intel TBB runtime library. This comparison was acheived by

applying both approaches on an industrial use case which consists in a co-simulation of a four cylinder spark ignition engine. The various tests that we performed showed the efficieny of our proposed heuristics. Second, we focused on the parallelization of FMI-based co-simulation under real-time constraints. In particular, we were interested in HiL (Hardware-in-the-Loop) co-simulation where a part of the co-simulation is replaced by its real counterpart that is physically available. The real and simulated parts have to exchange data during the execution of the co-simulation under real-time constraints. In other words, the inputs and ouputs of the real part are sampled periodically, sending and receving data to and from the simulated part. This periodic data exchange defines a set of real-time constraints to be satisfied by the simulated part. We proposed a method for defining these real-time constraints and propagating them to all the operations of the co-simulation (simulated part). In our ongoing work, we are focusing on multi-core scheduling of FMI-based co-simulation under real-time constraints. More precisely, we are working on a heuristic and an ILP-based algorithm that will enable the execution of the co-simulation on a multi-core processor while ensuring the defined real-time constraints are respected.

7.3. Principles of Probabilistic Composition

Participants: Slim Ben-Amor, Liliana Cucu, Adriana Gogonel, Cristian Maxim.

The statistical estimation of time parameters for real-time systems is proposed at two levels:

1. at program level and in this case we are dealing with timing analysis of programs that requires later appropriate probabilistic composition principles like reproducibility and representativity [3], [1]. For instance we have underlined in [14] the difficulties to ensure such properties for many-cores architectures.

   While we are proposing static analyses using worst-case bounds on the execution at instruction level for specialized architectures [2], we are interested also in proposing composition principles allowing to combine the timing impact of execution time variation factors, identified as a key open problem in the context of the timing analysis of programs while using the Extreme Value Theory [1]. Our composition solution is based on a Bayesian modeling that considers iteratively the inclusion of new factors while a representative measurement protocole is built [13] with respect to the reproducible Extreme Value Theory-based estimator that we have proposed.

2. at system level and in this case we are dealing with schedulability analysis of set of programs, a.k..a tasks, that requires appropriate composition principles like probabilistic independence while the dependence between tasks is taken into account. After proposing a first solution to the schedulability analysis of real-time probabilistic tasks in presence of precedence constraints on uniprocessor system [6], we explore the state of art of real-time scheduling on multiprocessor system and probabilistic real-time existing analysis. Our choice goes to partitioned multiprocessor scheduling to ensure the applicability of our previous results in the case of one processor. We have proposed a first optimal partitioning strategy based individual task utilization and we compare different tasks combinations that fit on a single processor following an utilization task ratio principle as partitioning choice. When assessing our method, a counter example of a possible optimality has appeared. Moreover this method has not an important improvement compared to existing partitioning strategies like best fit. Therefore we prepare the application of an existing solution to the bin packing problem [17] proposed in mathematics domain to partition real-time tasks on multiprocessor system in order to propose an appropriate probabilistic analysis.

   The exact schedulability analyses are often competing with statistical estimation of response time based on simulation and we propose such result in [9]. Such results allow to advance on the understanding of the notion of representativeness in the context of our problem that becomes today central in our community. The explosion of probabilistic schedulability analyses published in the last years have convinced us to join the book proposal of a Handbook on Real-Time Computing in order to integrate a comprehensive description of these analyses [4].

7.4. pWCET Estimation: a System Concern

Participants: Irina-Mariuca Asavoae, Mihail Asavoae, Slim Ben-Amor, Antoine Bertout, Liliana Cucu, Adriana Gogonel, Tomasz Kloda, Cristian Maxim, Walid Talaboulma.

From modelling to time validation, the design of an embedded system may benefit from a better utilisation of probabilities while providing means to prove their results. The arrival of new complex processors has made the time analysis of the programs more difficult while there is a growing need to integrate uncertainties from all levels of the embedded systems design. Probabilistic and statistical approaches are one possible solution and they require appropriate proofs in order to be accepted by both scientific community and industry. Such proofs cannot be limited at processor or program level and we plead for a system approach in order to take into account the possible interactions between different design levels by using the probabilistic formulation as compositional principle.

Our first arguments are provided by a valid statistical estimation of bounds on the execution time of a program on a processor. More precisely, the probabilistic worst-case execution time (pWCET) $\mathcal{C}$ of a program is an upper bound on all possible probabilistic execution times $\mathcal{C}_i$ for all possible execution scenarios $S_i, \forall i \geq 1$. According to EVT if the maximum of execution times of a program converges, then this maximum of the execution times $\mathcal{C}_i, \forall i \geq 1$ converges to one of the three possible Generalized Extreme Value (GEV) laws: Fréchet, Weibull and Gumbel corresponding to a shape parameter $\xi > 0$, $\xi < 0$, and $\xi = 0$, respectively. EVT has two different formulations: Generalized Extreme Value (GEV) and Generalized Pareto Distribution (GPD) and the difference between them is the way the extreme values are selected. GEV is based on the block maxima reasoning, grouping execution times by chronological groups (called blocks) and only the largest value of each group is considered as an extreme value. GPD is a method based on the threshold approach that considers only the values larger than the chosen threshold as extreme values. The voting procedure is based on the utilization of the both formulations of the EVT.

- **Block size estimation :** The GEV models obtained for different block sizes (BS), BS from 10 to $\frac{n}{10}$ are compared, where n is the cardinal of the trace of execution times. We compare the models fitting the extreme values corresponding to each choice of BS and the evolution of the shape parameter function of BS. We keep the BS that assures the best compromise between fitting the data and having a shape parameter within a stability interval of a range of shape parameters estimates. The way GEV models fit the data is analyzed within the tool by using a graphical method including the qqplot and the return level plot. We keep the GEV model corresponding to the shape parameter as the result of the aforementioned compromise and we compute the pWCET as the $1 - CDF$ (inverse of the cumulative distribution function) of the GEV.
- **Threshold level estimation :** The procedure is similar to the GEV procedure. All GPD models obtained for different threshold levels from 80% to 99%, are compared. In the same way as for GEV, we compare the models fitting the extreme values corresponding to each threshold and the evolution of the shape parameter function of threshold. At the end we keep the threshold level assuring the best compromise between fitting the data (graphical method) and having the shape parameter within a stability interval of a range of shape parameters estimates. We also consider the mean residual life plot (mean of excess) that may be consulted in case of a doubt between two different thresholds, we will prefer the threshold level such that the curve of mean of excess experiences linearity. We keep the GPD model corresponding to the shape parameter resulting from the aforementioned compromise and we compute the pWCET as the $1 - CDF$ of the GPD.
- **Comparing GEV and GPD pWCET estimates :** The comparison of the pWCET obtained with both methods, GEV and GPD is done graphically. Superposing the two curves allows to compare the distance between the two distributions. If an important difference is noticed, other GEV/GPD models are tested. In such cases calculating the pWCET estimate as a combination of GEV and GPD results is also recommended. A joint pWCET estimate is obtained by choosing for each probability the largest value between GEV and GPD . The tool implementing this method is available on line at `inria-rscript.serveftp.com` (requires a secured connection to be provided under request) [8].

- **Conditions of use :** The application of EVT requires to verify that the analyzed data are identically distributed, i.e., the execution times are following the same (unknown) probability distribution. That condition is tested before the analysis is started, and data is treated according to the test results. Another EVT applicability condition is the independence of the data. That condition is not mandatory in the sense that non-independent data can be analyzed. The case of dependent data can be split in two sub cases. The first one is where there are dependencies within the data, still the picked extremes values are independent. In that case the analysis will be done in the same way as for the independent data. The second case is the one where there are dependencies also between the extreme values. In that case one more step is added in the procedure. This step is the de-clustering process before applying GPD and the use of the index while GEV is applied.

During the second year of PhD thesis of Talaboulma Walid, we continued exploring solutions to WCET (Worst Case Execution Time) estimation and Real Time Scheduling on multiprocessors. WCET analysis done on a monoprocessor system (in isolation) can no longer be trusted to be accurate when we run our tasks on a multiprocessor (two processors), the problem of Co-runner interference arises and this is due to contention in shared hardware, two processors share the same memory and contention will occur when a simultaneous access is done, thus delaying one of the request, and this can counter-intuitively make programs run longer in a multiprocessor than what the analysis predicted on a monoprocessor, leading to deadline misses. In [20] authors evaluate explicit reservation of cache memory to reduce the cache-related preemption delay observed when tasks share a cache in a preemptive multitasking hard real-time system. Another solution is presented in [19] by management of tasks shared resources access using performance counter to stop tasks when they exceed their allocated budget (for instance cache misses) and thus providing guarantees on global memory bandwidths, moreover in [15] some offline analysis is done using heuristics to find optimal time triggered schedules for shared memory access.

We propose in our work to generate programs memory access profile, that we obtain by running tasks on a cycle accurate System Simulator, with a precise cycle accurate model of DDRAM memory controller and a full model of memory hierarchy including caches and main memory devices, and we log every memory event that occurs inside the simulation, our approach doesn't necessitate modifications of software layer, or recompilation of task code First we focus on simple tasks with few branches and simple memory access patterns as a proof of concept, and we choose a COTS (component of the shelf) platform with two complex processor cores. We intend to loosen those constraints when our analysis is matured. We use those profiles to account for co runners interference and add it to WCET value obtained in isolation, and then update our schedule, we can also insert idle times at correct scheduling events to decrease this interference, and in the future use a modified memory management system to pre-load specific memory areas into the cache and thus slide those access back in time to eliminate simultaneous memory access and converge toward an isolation WCET value.

7.5. Safe Parallelization of Hard Real-Time Avionics Software

Participants: Keryan Didier, Dumitru Potop-Butucaru.

This work took place in the framework of the ITEA3 ASSUME project, which funds the PhD thesis of Keryan Didier, and in close collaboration with Inria PARKAS, Airbus, and Kalray.

Concurrent programming is notoriously difficult, especially in constrained embedded contexts. Threads, in particular, are wildly nondeterministic as a model of computation, and difficult to analyze in the general case. Fortunately, it is often the case that multi-threaded, semaphore-synchronized embedded software implements high-level functional specifications written in a deterministic data-flow language such as Scade or (safe subsets of) Simulink.

In many cases, the multi-threaded implementation of such specifications preserves a fundamentally dataflow structure, with specific rules on the way platform resources (shared memory, semaphores) are used. When this happens, the implementation is best represented as a dataflow synchronous program whose elements are mapped on the platform resources. Ensuring the correctness of such an implementation consists in ensuring that:

1. The dataflow program (without the mapping) implements the semantics of the functional specification. This analysis can be performed inside the dataflow model.
2. Once the mapping of program elements onto the platform resources [0] is performed, the execution of the platform (under platform semantics) implements the behavior of the dataflow program.

Together, the dataflow program and the mapping information form an *implementation model*. This model is strictly richer than the multi-threaded C code, which can be obtained through a pretty-printing of model parts. Exposing the internal data-flow structure of the implementation facilitates defining and establishing correctness, *e.g.* the correctness of the synchronization or memory coherence protocols synthesized during the implementation process. All analyses can be realized using efficient tools specific to the synchronous model. Finally, if manual inspection of the C multi-threaded code is required, such a representation can be used to enforce strict code structuring rules which facilitate understanding.

We proposed a language for describing such implementation models that expose the data-flow behavior hiding under the form of a multi-threaded program. The language allows the representation of efficient implementations featuring pipelined scheduling and optimized memory allocation and synchronization [12].

We also proposed a design and tool flow taking as input industrial specifications based on Lustre/Scade and automatically producing fully mapped parallel implementation models and implementations with hard real-time guarantees. The front-end of the flow implements properties facilitating the mapping, e.g., exposing the state of all nodes to memory optimization. To strictly enforce realtime guarantees, the offline mapping algorithms of the back-end consider all sources of interference, including concurrent memory accesses, coherence protocols and event-driven synchronization. Our flow scales to an avionics application comprising more than 5000 unique nodes, targeting the Kalray MPPA 256 many-core platform, selected for its timing predictability.

7.6. Real-time Platform Modeling

Participants: Fatma Jebali, Dumitru Potop-Butucaru.

One key difficulty in embedded systems design is related to the existence of multiple models of the same system, at different abstraction levels, and used in various phases of the design flow. Usual models include *cycle-accurate, bit-accurate (CABA)* system models used to perform exact simulation for precision tuning, microarchitectural models used during WCET (*Worst-Case Execution Time*) analysis of sequential tasks, and high-level models used during WCRT (*Worst-Case Response Time*) analysis of the whole system. In current practice, these models are developed separately, and it is difficult to ensure (by extensive simulation) that they are consistent.

We explore the possibility of obtaining both a CABA and a WCET microarchitectural simulator from a single source, along with a formal consistency guarantee. This year we considered the timing abstraction issue: Both CABA and WCET simulators use a cycle-based execution model, but the cycle corresponds in one case to hardware clock cycles, and in the other to PC (program counter) advancement. We showed that for architectures satisfying a scheduling-independence property (known as in-order architectures) it is possible to produce from a single source both types of simulations (clock-driven and PC-driven), with a formal correctness guarantee. Preliminary results have been presented at the Synchron'07 workshop.

[0] Sequencing of blocks into threads executed by processors; code, stack and data variables to memory locations; synchronizations to semaphores, *etc.*

HYCOMES Project-Team

5. New Results

5.1. Semantics, Static or Runtime Analysis of Hybrid Systems

5.1.1. Structural Analysis of Multi-Mode DAEs

Differential Algebraic Equation (DAE) systems constitute the mathematical model supporting physical modeling languages such as Modelica or Simscape. Unlike Ordinary Differential Equations, or ODEs, they exhibit subtle issues because of their implicit *latent equations* and related *differentiation index*. Multi-mode DAE (mDAE) systems are much harder to deal with, not only because of their mode-dependent dynamics, but essentially because of the events and resets occurring at mode transitions. Unfortunately, the large literature devoted to the numerical analysis of DAEs do not cover the multi-mode case. It typically says nothing about mode changes. This lack of foundations cause numerous difficulties to the existing modeling tools. Some models are well handled, others are not, with no clear boundary between the two classes. In [11], we develop a comprehensive mathematical approach to the *structural analysis* of mDAE systems which properly extends the usual analysis of DAE systems. We define a constructive semantics based on nonstandard analysis and show how to produce execution schemes in a systematic way. This work has been accepted for presentation at the HSCC 2017 conference [18] in April 2017.

5.1.2. Operational Models for Piecewise-Smooth Systems

In [7], we study ways of constructing meaningful operational models of piecewise-smooth systems (PWS). The systems we consider are described by polynomial vector fields defined on non-overlapping semi-algebraic sets, which form a partition of the state space. Our approach is to give meaning to motion in systems of this type by automatically synthesizing operational models in the form of hybrid automata (HA). Despite appearances, it is in practice often difficult to arrive at satisfactory HA models of PWS. The different ways of building operational models that we explore in our approach can be thought of as defining different semantics for the underlying PWS. These differences have a number of interesting nuances related to phenomena such as chattering, non-determinism, so-called mythical modes and sliding behaviour.

5.1.3. Accelerated Simulation of Hybrid Systems: Method combining static analysis and runtime execution analysis

Ayman Aljarbouh has defended his PhD [4] on September 13th 2017. His PhD has been partially funded by an ARED grant of the Brittany Regional Council. His doctoral work took place in the context of the Modrio (completed in 2016) and Sys2Soft (completed in 2015) projects on hybrid systems modeling. Ayman Aljarbouh has been working on accelerated simulation techniques for hybrid systems. In particular, he has contributed, and implemented in a software prototype, a regularisation method transforming automatically at runtime a chattering behaviour into a semantics preserving smooth behaviour. He has also contributed a method for the approximation of Zeno behaviour. This method enables to jump past an accumulation of an infinite number of zero-crossing events, and to continue the simulation of a large class of Zeno hybrid systems, after accumulation points.

5.1.4. A Type-based Analysis of Causality Loops in Hybrid Systems Modelers

Explicit hybrid systems modelers like Simulink/Stateflow allow for programming both discrete- and continuous-time behaviors with complex interactions between them. A key issue in their compilation is the static detection of algebraic or causality loops. Such loops can cause simulations to deadlock and prevent the generation of statically scheduled code. In [5], we addresses this issue for a hybrid modeling language that combines synchronous data-flow equations with Ordinary Differential Equations (ODEs). We introduce the operator last(x) for the left-limit of a signal x. This operator is used to break causality loops and permits a

uniform treatment of discrete and continuous state variables. The semantics relies on non-standard analysis, defining an execution as a sequence of infinitesimally small steps. A signal is deemed causally correct when it can be computed sequentially and only changes infinitesimally outside of announced discrete events like zero-crossings. The causality analysis takes the form of a type system that expresses dependences between signals. In well-typed programs, signals are provably continuous during integration provided that imported external functions are also continuous. The effectiveness of this system is illustrated with several examples written in Zélus, a Lustre-like synchronous language extended with hierarchical automata and ODEs.

5.2. Formal Verification of Hybrid Systems

5.2.1. Formal Verification of Station Keeping Maneuvers for a Planar Autonomous Hybrid System

In [9], we investigate the formal verification of a hybrid control law designed to perform a station keeping maneuver for a planar vehicle. Such maneuver requires that the vehicle reaches a neighborhood of its station in finite time and remains in it while waiting for further commands. We model the dynamics as well as the control law as a hybrid program and formally verify the reachability and safety properties involved. We highlight in particular the automated generation of invariant regions which turns out to be crucial in performing such verification. We use the hybrid system theorem prover KeymaeraX to formally check the parts of the proof that can be automatized in the current state of the tool.

5.2.2. Formal verification of obstacle avoidance and navigation of ground robots

In [6], we answer fundamental safety questions for ground robot navigation: Under which circumstances does a given control decision make a ground robot safely avoid obstacles? Unsurprisingly, the answer depends on the exact formulation of the safety objective as well as the physical capabilities and limitations of the robot and the obstacles. Because uncertainties about the exact future behavior of a robot's environment make this a challenging problem, we formally verify corresponding controllers and provide rigorous safety proofs justifying why they can never collide with the obstacle in the respective physical model. To account for ground robots in which different physical phenomena are important, we analyze a series of increasingly strong properties of controllers for increasingly rich dynamics and identify the impact that the additional model parameters have on the required safety margins. We analyze and formally verify: (i) static safety, which ensures that no collisions can happen with stationary obstacles, (ii) passive safety, which ensures that no collisions can happen with stationary or moving obstacles while the robot moves, (iii) the stronger passive friendly safety in which the robot further maintains sufficient maneuvering distance for obstacles to avoid collision as well, and (iv) passive orientation safety, which allows for imperfect sensor coverage of the robot, i. e., the robot is aware that not everything in its environment will be visible. We formally prove that safety can be guaranteed despite sensor uncertainty and actuator perturbation. We complement these provably correct safety properties with liveness properties: we prove that provably safe motion is flexible enough to let the robot navigate waypoints and pass intersections. In order to account for the mixed influence of discrete control decisions and the continuous physical motion of the ground robot, we develop corresponding hybrid system models and use differential dynamic logic theorem proving techniques to formally verify their correctness. Since these models identify a broad range of conditions under which control decisions are provably safe, our results apply to any control algorithm for ground robots with the same dynamics. As a demonstration, we, thus, also synthesize provably correct runtime monitor conditions that check the compliance of any control algorithm with the verified control decisions.

5.3. Synchronous Interfaces and Assume/Guarantee Contracts

In [10], we establish a link between the theory of Moore Interfaces proposed in 2002 by Chakraborty et al. as a specification framework for synchronous transition systems, and the Assume/Guarantee contracts as proposed in 2007 by Benveniste et al. as a simple and flexible contract framework. As our main result we show that the operation of saturation of A/G contracts (namely the mapping $(A, G) \rightarrow (A, G \vee \neg A)$), which was considered a drawback of this theory, is indeed implemented by the Moore Game of Chakraborty et al. We further develop this link and come up with some remarks on Moore Interfaces.

5.4. CominWeb project of the Labex CominLabs

Jean Hany and Albert Benveniste (together with William Dedzoe) were involved in this project.

CominWeb is a project supported by the Labex CominLabs since 2013. Its original objective was to equip CominLabs with Web infrastructures, tools, and services, that would allow to run the scientific activity of the Labex in an innovative way. Based on a study of the population of the CominLabs researchers, performed in year 2014-15 by the teams of CominLabs involved in social sciences, several services were investigated and prototyped. A short trial addressed the automatic generation of a scientific activity report, for a CominLabs project, from the material available from the publications ot the project team. This was suspended because such a service was not considered very useful by the community. A second trial (nicknamed "NSA") consisted in monitoring the flows of email exchanges addressed to aliases of the CominLabs projects, with the objective of classifying the mails into: meeting announcements, mails with attachments of interest, and other mails. This would give to the CominLabs head a view on the project's activities without asking for any specific contribution from the researchers. This was more interesting. Still, a difficulty was that researchers did not use the project aliases so much. For priority issues, this development was also suspended.

The main result of this project is thus the service called *LookinLabs*, deployed in two different versions: http://lookinlabs4halinria.cominlabs.ueb.eu/ and http://www.lookinlabs.cominlabs.ueb.eu/. The former is a more advanced version of LookinLabs, developed for the whole Inria community, by exploiting the HAL publication archive. LookinLabs for HAL-Inria allows the user to find, among teams/individuals/publications taken from all the Inria teams, those best matching a query consisting of a list of keywords or a short text. The tool exploits, as data, HAL-Inria archives, in combination with the Inria Activity reports (the Raweb), and the internal data base of Inria teams called BASTRI. Active teams/individuals are shown in boldface. Teams/individuals shown in gray are no longer active at Inria. If team TEAM0 is no longer active, the mention: TEAM0 $\rightarrow$ (TEAM1,TEAM2) indicates follow-up active teams, if any. In LookinLabs, no ontology is used. No data need to be manually entered (besides the users' queries). The tool uses *Elasticsearch* (https://www.elastic.co/fr/products/elasticsearch) as its core algorithm. This means that the matching is based on a distance between the query and the set of data attached, in HAL, to each team/individual/publication. Ranking is performed accordingly. Explanations are given for each returned item. Correlation graphs are given, allowing to navigate through teams or individuals that share common interests (they may or may not be co-authors).

LookinLabs is deployed in two versions. LookinLabs4HALInria is the one we just described. The other version is in operation since 2016 and addresses the scientific community of CominLabs researchers. The data used are up to 10 standard bibliographical data bases (Dblp, IEEE Explore, Arxiv, HAL, and more) for which links have been collected from the researchers (this was the only data they were asked for). Results are returned in the form of individuals and publications, not teams.

KAIROS Team

6. New Results

6.1. Formal Model-Based, Platform-Based System Engineering for CPS heterogeneous systems

Participants: Julien Deantoni, Robert de Simone, Frédéric Mallet, Marie-Agnès Peraldi-Frati, Ales Mishchenko.

The proper inclusion of our models and techniques into a realistic or proto-industrial design flow is a topic of discussion with partners, which brings back a number of fundamental issues about the nature of modeling itself, specially to allow and promote further analysis. The modeling of execution platforms, which may seem primarily architectural, is itself in fact subject to combined physical model interplay with functionality (as in power consumption and heat dissipation in processors); Internet of Things and industrial embedded systems are new specific focuses for us here. In this context we have been concentrating in the course of the PIA Clarity project (see 7.1) on the CAPELLA system engineering language, and its multi-view aspects; in the new IRT Saint-Exupery ATIPPIC collaboration we intend to focus on a realistic-scale use-case modeled with CAPELLA and proper formal annotations for the design of micro-satellite systems with COTS processor architectures.

6.2. Cyber extensions for the FMI physical API

Participants: Julien Deantoni, Giovanni Liboni, Robert de Simone.

The challenge here is to specify how an appropriate behavioral interface can be specified at the language level so that individual simulators can provide sufficient interface information to allow their correct coordination in terms of correctness and performance. Using our modeling approaches to describe formally the specificities of simulators and scheduling patterns, maybe even targeting synthesis of efficient interactions, is a new research topic [9], conducted in art in the context of the GLOSE project (see 7.1.2) in collaboration with SAFRAN tech.

6.3. Logical Time and Uncertain Physical environments

Participants: Frédéric Mallet, Robert de Simone, Dongdong An.

We developed a specific formalism to express logical time constraints on models. Currently it remains mainly aimed at purely “cyber” models. Combination with continuous physical models, and stochastic uncertainty, will be further studied at the level of constraint expressiveness. The issue of expressing meaningful discrete clocks from stochastic clock schemes, and combining the resulting clocks borne from physics to other clocks borne from sampling oberver environment, is a topic of collaboration with ECNU Shanghai through the PhD long-term visit of AN Dongdong to Sophia, and other discusions as well [3].

6.4. Automatic analysis and verification for specific classes of models

Participants: Robert de Simone, Frédéric Mallet, Emilien Kofman, Julien Deantoni.

This part aims at continuing and extending our work on “mostly-automatic”verification, with SMT solvers and model-checking in place to compute optimal scheduling and allocation, for instance. The new expressiveness of CPS adds new challenges [10]. While the global problem of schedulability for a generic CCSL specification is an open problem. We have identified a particular subclass of scheduling called "periodic scheduling" and establish a sufficient conditions for detecting if there is a valid periodic scheduling for a given CCSL specification. This condition is checked using the MAUDE rewriting system that directly encodes the operational semantics of CCSL [8]. However, the performances highly depend on the "period" of the periodic scheduling.

A more efficient solution based on SMT solver is under investigation and has given first encouraging results. A specific direction is reported in Emilien Kofman PhD thesis [1].

6.5. Behavioral semantics and equivalence notions for Open Systems

Participants: Eric Madelaine, Xudong Qin.

Model-Based Design naturally implies model transformations. To be proven correct, they require equivalence of "Open" terms, in which some individual component models may be omitted. Proper behavioral equivalence in the domain of CPS (which may include variants) is the challenge here. Such models take into account various kind of data parameters, including, but not limited to, time. The middle term goal is to build a formal framework, but also an effective tool set, for the compositional analysis of such programs. Our joint work (between Eric Madelaine and Ludovic Henrio from France, professors Zhang Min, Deng Yuxin and students from ECNU) on symbolic approaches to the composition of concurrent processes has progressed mainly on the pratical side, with an implementation of a prototype algorithm computing the symbolic semantics (called Open Automata) of open systems, and validating the approach for encoding constructs of various formalisms. A paper has been submitted for publication. As a particular set of use-cases, we have started using pNets to encode the behavior of "Architecture Templates" of the BIP language, with the aim of proving generic properties of these constructs, and building full systems by combining such architectures, with proven guarantees.

We have published preliminary work proposing a framework for open systems defining their symbolic semantics and some verification mechanisms (equivalences, model-checking), and we have started developping prototype tools supporting this approach.

This is joint work with ZHANG Min, from ECNU Shanghai, partially conducted in the framework of the FM4CPS associated-team.

6.6. Formal model of computations, mobility and resource discovery adapted for CPS

Participants: Robert de Simone, Luigi Liquori.

We will adapt formal Models of Computations and formal system like Type Theory to capture properties for distributed, networked and mobile CPS. Discovering and synchronizing, possibly in presence of mobility, are important issues in CPS. We will extend our past distributed resource discovery model with a notion of logical time; the difficulty comes to both discover and synchronize CPSs having different logical times.

6.7. Logical Frameworks with logical time

Participants: Luigi Liquori, Robert de Simone.

Adapting the Logical Framework (LF) based on Type Theory to better understand the analysis and verification of CPS is the challenge here. Previous works on extending Logical Framework with Locked Types and Proof-functional logics could captures the notion synchronizing (via a suitable constraint resolution) two proof systems: the internal and the external one; this type can be viewed as a form of communication between the internal and the external logic. Applications could be feed by suitable outputs of the CPS co-modeling pillar phase like, e.g. CCSL expressions.

6.8. From code to model

Participant: Sid Touati.

This research result is about modeling code performances using gaussian mixtures. Thanks to this performance model, we propose additional precise performance metrics. Our additional statistical metrics for analysing and comparing program performances give the user more precise decision tools to select best code versions, not necessarily based on mean or median numbers. Also, we provide a new metric to estimate performance variability based on Gaussian mixture model. Our statistical methods are implemented with R and distributed as open source code.

This is a collaboration with Julien Worms (LMV, UVSQ), reported in [7].

6.9. From model to code

Participants: Amine Oueslati, Robert de Simone, Arsak Megkrampian.

Synthesizing programs and algorithms in an optimal fashion may call for model transformations, as above, but also on careful tuning at program level, which make take into consideration a number of concrete phenomena absent from abstract models (at the cost of losing exhaustivity). Specific code optimization techniques in a CPS design context is the challenge here. WE are collaborating with Dumitru Potop-Butucaru (formerly from the Aoste team) on the topic, where he develops the LoPhT (Logical to Physical Time) environment dealing with these issues.

PARKAS Project-Team

6. New Results

6.1. Compiler Optimisations for Multicore Architectures

Participants: Robin Morisset, Francesco Zappa Nardelli.

Robin has completed his research work on sound optimisations for modern multicore architectures. This covered optimisations that can be expressed inside the semantics of the C11/C++11 programming language, as well as optimisations that can be expressed only at the harware level. In particular we have shown how partial redundancy elimination (PRE) can be instantiated to perform *provably correct* fence elimination for multi-threaded programs running on top of the x86, ARM and IBM Power relaxed memory models. We have implemented our algorithm in the x86, ARM and Power backends of the `LLVM` compiler infrastructure. The optimisation does not induce an observable overhead at compile-time and can result in up-to 10% speedup on some benchmarks.

This work has been published in CC 2017 [10]. The implementation of the optimisations will be submitted for inclusion in the LLVM compiler suite.

Robin Morisset completed this line of research and defended his PhD Thesis in April 2017.

6.2. Julia Subtyping Reconstructed

Participant: Francesco Zappa Nardelli.

Julia is a programming language recently designed at MIT to support the needs of the scientific community. Julia occupies a unique position in the design landscape, it is a dynamic language with no type system, yet it has a surprisingly rich set of types and type annotations used to specify multimethod dispatch. The types that can be expressed in function signatures include parametric union types, covariant tuple types, parametric user-defined types with single inheritance, invariant type application, and finally types and values can be reified to appear in signatures. With Vitek started a research project to study the design and the pragmatic use of the Julia language. At first we focused on the Julia subtyping algorithm. We studied the empirical evidence that users appeal to all the features provided by Julia and we report on a formalisation and implementation of the subtyping algorithm. The work on subtyping is under submission to an international conference. This line of research will be pursued in the next year, studying method dispatch and type inference.

6.3. Comparing Designs for Gradual Types

Participant: Francesco Zappa Nardelli.

The enduring popularity of dynamically typed languages has given rise to a cottage industry of static type systems, often called gradual type systems, that let developers annotate legacy code piecemeal. Type soundness for a program which mixes typed and untyped code does not ensure the absence of errors at runtime, rather it means that some errors will caught at type checking time, while other will be caught as the program executes. After a decade of research it is clear that the combination of mutable state, self references and subtyping presents interesting challenges to designers of gradual type systems. We have reviewed the state of the art in gradual typing for objects, and introduced a class-based object calculus with a static type system, dynamic method dispatch, transparent wrappers and dynamic class generation that we use to model key features of several gradual type systems by translation to it, and discuss the implications of the respective designs. We have submitted this work to an international conference.

6.4. Symbolic Simulation for a timed-automaton subset of Zélus

Participants: Guillaume Baudart, Timothy Bourke, Marc Pouzet.

Synchronous languages like Lustre are ideal for programming an important class of embedded controllers. Their discrete model of time and deterministic semantics facilitate the precise expression of reactive behaviors. That said, many systems are naturally modeled using physical timing constraints that almost inevitably involve some 'timing nondeterminism' due to tolerances in requirements or uncertainties in implementations. Conversely, such constraints are readily modeled using Timed Automata, and simulated symbolically in Uppaal, but large-scale discrete-time behaviors are more cumbersome to express in such tools.

In this work, we combined existing techniques and data structures for Timed Safety Automata with typing and compilation techniques for synchronous languages to develop a novel programming language where discrete reactive logic can be mixed with nondeterministic continuous-time features. In particular, we developed an extension of Lustre and a specialization of Zélus for modeling real-time reactive systems, proposed a symbolic simulation scheme based on 'sweeping', and showed how to implement it via source-to-source compilation. A type system, based on that of Zélus, ensures the correct composition of discrete-time and continuous-time elements.

Our proposal has been implemented using the Zélus compiler and a small library of operations on Difference-Bound Matrices (DBMs). Unlike the work around Uppaal, we do not address verification or treat industrial case studies. A future direction could be to verify programs in our 'extended version of Lustre' by either generating C code and using the highly-tuned Uppaal DBM library, or combining symbolic techniques for Lustre programs with those for Timed Automata.

This work was presented at FDL 2017 [5]. A prototype implementation is available online.

This work is also described with extended examples in Baudart's PhD thesis [1] which was defended in March of 2017.

6.5. Verified compilation of Lustre

Participants: Timothy Bourke, Lélio Brun, Marc Pouzet.

Synchronous dataflow languages and their compilers are increasingly used to develop safety-critical applications, like fly-by-wire controllers in aircraft and monitoring software for power plants. A striking example is the SCADE Suite tool of ANSYS/Esterel Technologies which is DO-178B/C qualified for the aerospace and defense industries. This tool allows engineers to develop and validate systems at the level of abstract block diagrams that are automatically compiled into executable code.

Formal modelling and verification in an interactive theorem prover can potentially complement the industrial certification of such tools to give very precise definitions of language features and increased confidence in their correct compilation; ideally, right down to the binary code that actually executes.

This year we continued work on our verified Lustre compiler. We developed a set of benchmarks and evaluated the Worst Case Execution time of code generated by our compiler with that of code generated by the academic Heptagon and Lustre v6 compilers. This work also required numerous improvements to the parser and elaborator. We also tested the compiler on an industrial example in the context of the ASSUME project. We completed the end-to-end theorem showing that the dataflow semantics of input programs is preserved by the assembly language semantics generated by our compiler combined with the CompCert compiler. This work was presented in June at PLDI [8].

In the latter half to the year we worked on extending the compiler to accept nodes with clocked arguments, treating non-normalized Lustre, and adding a modular reset to the language.

To accept clocked arguments, we extended the semantic model, developed a richer encoding of the clock system, added a new invariant to forbid non-trivial sub-clocked expressions, and adapted the correctness proof. An unexpected complication was the need to pass undefined variables in function call arguments: this required changes to our intermediate Obc language and introduces minor technical difficulties in the translation to Clight which requires that variables be defined. This work is now almost complete.

To treat non-normalized Lustre, we introduced new syntactic and semantic definitions, updated the parser, and complete reworked the elaboration and type-checking passes. We developed many small Lustre programs to confirm our understanding of the language and test the updated front-end; this also revealed several bugs in other academic Lustre compilers. This work is now complete. The next step is to implement the normalization pass to connect the new front-end to the existing compilation passes.

The work on modular resets continues as part of L. Brun's PhD thesis. This year we developed a novel semantic model for modular resets and started considering how to generate provably correct code.

In collaboration with Pierre-Évariste Dagand (CNRS), Lionel Reig (Collège de France), and Xavier Leroy (Inria, GALLIUM team).

6.6. Zélus: Synchronous Languages + Ordinary Differential Equations

Participants: Timothy Bourke, Marc Pouzet.

Zélus is a synchronous language extended with Ordinary Differential Equations (ODEs) to model systems with complex interactions between discrete-time and continuous-time dynamics. It shares the basic principles of Lustre with features from Lucid Synchrone (type inference, hierarchical automata, and signals). The compiler is written in OCaml and is structured as a series of source-to-source and traceable transformations that ultimately yield statically scheduled sequential code. Continuous components are simulated using off-the-shelf numerical solvers (here Sundials CVODE) and, for the moment, two built-in solvers (ode23 and ode45).

Zélus is used to experiment with new techniques for building hybrid modelers like Simulink/Stateflow and Modelica on top of a synchronous language. The language exploits novel techniques for defining the semantics of hybrid modelers, it provides dedicated type systems to ensure the absence of discontinuities during integration and the generation of sequential code. In particular, all discrete computations must be aligned to zero-crossing events; programs with causality loops and uninitialized values are statically rejected.

This year we added arrays with iterators and statically expanded higher-order functions to the language. Both extensions required adapting the existing type and causality systems, and extending the compilation algorithms. These extensions allowed us to show that a fairly large set of blocks from the Simulink standard library can be programmed in a precise, purely functional language using stream equations, hierarchical automata, Ordinary Differential Equations (ODEs), and deterministic synchronous parallel composition. Although some blocks cannot be expressed as they mix discrete-time and continuous-time signals in unprincipled ways; they are statically rejected by the type checker. This work was presented at EMSOFT in October [9]

Our work on analyzing causality loops in hybrid systems modelers was published in the NAHS journal [2].

In collaboration with B. Caillaud and A. Benveniste (Inria Rennes); and F. Carcenac, B. Pagano, and C. Pasteur (ANSYS/Esterel Technologies).

6.7. Compiling synchronous languages for multi-processor implementations

Participants: Timothy Bourke, Albert Cohen, Guillaume Iooss, Marc Pouzet.

Working together with industrial partners in the context of the ASSUME project.

We spent a week in Toulouse working at Airbus on their use case and our front-end tools. We can now treat the case and generate code for Lopht (AOSTE team), which, in turn, generates executable code for the Kalray MPPA. We have also advanced significantly on two use cases provided by Safran. The first one is similar to the Airbus use case. The second one is more preliminary, it revealed the need for more general iterators to better express FFT algorithms.

We have made solid progress on a language extension for expressing and manipulating harmonic clocks. In particular, we derive a scheduling problem from the clock constraints in a program and we are working on automatically calculating their initial phases.

We have written an import tool that transforms graphs of dependencies between several Lustre components scheduled with different harmonic periods into a monolithic Lustre program. We are working on a hyper-scheduling transformation that generates a single step function running at the slowest period and that contains multiple instances of the faster tasks with annotations to ensure they execute at the correct time.

In collaboration (this year) with Dumitru Potop-Butucaru and Keryan Didier (Inria, AOSTE team); Jean Souyris and Adrien Gauffriau (Airbus); Philippe Baufreton et Jean-Marie Courtelle (Safran).

SPADES Project-Team

6. New Results

6.1. Components and contracts

Participants: Alain Girault, Christophe Prévot, Sophie Quinton, Jean-Bernard Stefani.

6.1.1. Contracts for the negotiation of embedded software updates

We address the issue of change during design and after deployment in safety-critical embedded system applications, in collaboration with Thales and also in the context of the CCC project (http://ccc-project.org/).

In collaboration with Thales, we mostly focus on timing aspects with the objective to anticipate, at design time, future software evolutions and identify potential schedulability bottlenecks. This year we have paved the way for an extension, to more complex systems, of the approach developed last year to quantify the flexibility of a system with respect to timing. Specifically, we have focused on systems with task chains, and have proposed new methods for computing upper and lower bounds on task chain latencies. This work will be submitted to a conference early 2018. Our methods are also being implemented in the Thales tool chain, in order to be used in industry.

6.1.2. Location graphs

The design of configurable systems can be streamlined and made more systematic by adopting a component-based structure, as demonstrated with the FRACTAL component model [38]. However, the formal foundations for configurable component-based systems, featuring higher-order capabilities where components can be dynamically instantiated and passivated, and non-hierarchical structures where components can be contained in different composites at the same time, are still an open topic. We have recently introduced the location graph model [70], where components are understood as graphs of locations hosting higher-order processes, and where component structures can be arbitrary graphs.

We have continued the development of location graphs, revisiting the underlying structural model (hypergraphs instead of graphs), and simplifying its operational semantics while preserving the model expressivity. Towards the development of a behavioral theory of location graphs, we have defined different notions of bisimilarity for location graphs and shown them to be congruences, although a fully fledged co-inductive characterization of contextual equivalence for location graphs is still in the works. This work has not yet been published.

6.2. Real-Time multicore programming

Participants: Pascal Fradet, Alain Girault, Gregor Goessler, Xavier Nicollin, Sophie Quinton.

6.2.1. Dynamicity in dataflow models

Recent dataflow programming environments support applications whose behavior is characterized by dynamic variations in resource requirements. The high expressive power of the underlying models (*e.g.*, Kahn Process Networks or the CAL actor language) makes it challenging to ensure predictable behavior. In particular, checking *liveness* (*i.e.*, no part of the system will deadlock) and *boundedness* (*i.e.*, the system can be executed in finite memory) is known to be hard or even undecidable for such models. This situation is troublesome for the design of high-quality embedded systems. In the past few years, we have proposed several parametric dataflow models of computation.

We have written a survey that provides a comprehensive description of the existing parametric dataflow MoCs (constructs, constraints, properties, static analyses) and compares them using a common example [10]. The main objectives are to help designers of streaming applications to choose the most suitable model for their needs and to pave the way for the design of new parametric MoCs.

We have studied *symbolic* analyses of dataflow graphs [11]. Symbolic analyses express the system performance as a function of parameters (*i.e.*, input and output rates, execution times). Such functions can be quickly evaluated for each different configuration or checked *w.r.t.* different quality-of-service requirements. These analyses are useful for parametric MoCs, partially specified graphs, and even for completely static SDF graphs. Our analyses compute the maximal throughput of acyclic synchronous dataflow graphs, the minimum required buffers for which as soon as possible (asap) scheduling achieves this throughput, and finally the corresponding input-output latency of the graph.

We have proposed an original method to deal with lossy communication channels in dataflow graphs. Lossy channels intrinsically violate the dataflow model of computation. Yet, many real-life applications encounter some form of lossy channels, for instance IoT applications. The challenge that is raised is how to manage the retransmissions in case of lost or corrupted tokens. The solution that we have proposed involves decomposing the execution of the dataflow graph into three phases: (i) an upstream phase where all the actors before the lossy channel are executed as usual; (ii) a lossy phase where only the two actors linked by the lossy channel are executed, as many times as required until all the tokens are correctly transmitted; and (iii) a downstream phase where all the actors after the lossy channel are executed as usual. When a graph includes several lossy channels, things become more complex. We rely on the Boolean parameters of BPDF [32] to encode enabling conditions on channels so that the execution follows this upstream-lossy-downstream semantics [12].

We are now studying models allowing dynamic reconfigurations of the *topology* of the dataflow graphs. This would be of interest for C-RAN and 5G telecommunication applications. This is one of the research topic of Arash Shafiei's PhD in collaboration with Orange Labs.

6.2.2. Synthesis of switching controllers using approximately bisimilar multiscale abstractions

The use of discrete abstractions for continuous dynamics has become standard in hybrid systems design (see *e.g.*, [71] and the references therein). The main advantage of this approach is that it offers the possibility to leverage controller synthesis techniques developed in the areas of supervisory control of discrete-event systems [66]. The first attempts to compute discrete abstractions for hybrid systems were based on traditional systems behavioral relationships such as simulation or bisimulation, initially proposed for discrete systems most notably in the area of formal methods. These notions require inclusion or equivalence of observed behaviors which is often too restrictive when dealing with systems observed over metric spaces. For such systems, a more natural abstraction requirement is to ask for closeness of observed behaviors. This leads to the notions of approximate simulation and bisimulation introduced in [45].

These approaches are based on sampling of time and space where the sampling parameters must satisfy some relation in order to obtain abstractions of a prescribed precision. In particular, the smaller the time sampling parameter, the finer the lattice used for approximating the state-space; this may result in abstractions with a very large number of states when the sampling period is small. However, there are a number of applications where sampling has to be fast; though this is generally necessary only on a small part of the state-space. We have been exploring two approaches to overcome this state-space explosion [4].

We are currently investigating an approach using mode sequences of given length as symbolic states for our abstractions. By using mode sequences of variable length we are able to adapt the granularity of our abstraction to the dynamics of the system, so as to automatically trade off precision against controllability of the abstract states.

6.2.3. Schedulability of weakly-hard real-time systems

We focus on the problem of computing tight deadline miss models for real-time systems, which bound the number of potential deadline misses in a given sequence of activations of a task. In practical applications, such guarantees are often sufficient because many systems are in fact not hard real-time [3].

We have developed an extension of sensitivity analysis for budgeting in the design of weakly-hard real-time systems [18]. During design, it often happens that some parts of a task set are fully specified while other parameters, *e.g.*, regarding recovery or monitoring tasks, will be available only much later. In such cases, sensitivity analysis can help anticipate how these missing parameters can influence the behavior of the whole

system so that a resource budget can be allocated to them. We have developed an extension of sensitivity analysis for deriving task budgets for systems with hard and weakly-hard requirements. This approach has been validated on synthetic test cases and a realistic case study given by our partner Thales.

A second contribution in this area is the application of our method for computing deadline miss models, called Typical Worst-Case Analysis (TWCA), to systems with finite queue capacity [9]. Finite ready queues, implemented by buffers, are a system reality in embedded real-time computing systems and networks. The dimensioning of queues is subject to constraints in industrial practice, and often the queue capacity is sufficient for typical system behavior, but is not sufficient in peak overload conditions. This may lead to overflow and consequently to the discarding of jobs. In this paper, we explore whether finite queue capacity can also be used as a mean of design in order to reduce workload peaks and thus shorten a transient overload phase. We have proposed an analysis method which is to the best of our knowledge the first one able to give (a) worst-case response times guarantees as well as (b) weakly-hard guarantees for tasks which are executed on a computing system with finite queues. Experimental results show that finite queue capacity may only have weak overload limiting effect. This unexpected outcome can be explained by the system behavior in the worst-case corner cases. The analysis shows nevertheless that a trade-off between weakly-hard guarantees and queue sizes is possible.

Finally, in collaboration with TU Braunschweig and Daimler we have worked on the application of the Logical Execution Time (LET) paradigm, according to which data are read and written at predefined time instants, to the automotive industry. Specifically, we have bridged the gap between LET, as it was originally proposed [59], and its current use in the automotive industry. One interesting outcome of this research is that it can nicely be combined with the use of TWCA. This work has not been published yet.

6.2.4. A Markov Decision Process approach for energy minimization policies

In the context of independent real-time sporadic jobs running on a single-core processor equipped with Dynamic Voltage and Frequency Scaling (DVFS), we have proposed a Markov Decision Process approach (MDP) to compute the scheduling policy that dynamically chooses the voltage and frequency level of the processor such that each job meets its deadline and the total energy consumption is minimized. We distinguish two cases: the finite case (there is a fixed time horizon) and the infinite case. In the finite case, several *offline* solutions exist, which all use the complete knowledge of all the jobs that will arrive within the time horizon [74], *i.e.*, their size and deadlines. But clearly this is unrealistic in the embedded context where the characteristics of the jobs are not known in advance. Then, an optimal offline policy called Optimal Available (OA) has been proposed in [30]. Our goal was to improve this result by taking into account the *statistical characteristics* of the upcoming jobs. When such information is available (for instance by profiling the jobs based on execution traces), we have proposed several speed policies that optimize the *expected* energy consumption. We have shown that this general constrained optimization problem can be modeled as an unconstrained MDP by choosing a proper state space that also encodes the constraints of the problem. In particular, this implies that the optimal speed at each time can be computed using a *dynamic programming* algorithm, and that the optimal speed at any time t will be a deterministic function of the current state at time t [21]. This is the topic of Stephan Plassart's PhD, funded by the CASERM Persyval project.

6.2.5. Formal proofs for schedulability analysis of real-time systems

We have started to lay the foundations for computer-assisted formal verification of schedulability analysis results. Specifically, we contribute to Prosa [26], a foundational Coq library of reusable concepts and proofs for real-time schedulability analysis. A key scientific challenge is to achieve a modular structure of proofs for response time analysis. We intend to use this library for:

1. a better understanding of the role played by some assumptions in existing proofs;
2. a formal comparison of different analysis techniques; and
3. the verification of proof certificates generated by instrumenting (existing and efficient) analysis tools.

Two schedulability analyses for uniprocessor systems have been formalized and mechanically verified in Coq for:

- sporadic task sets scheduled according to the Time Division Multiple Access (TDMA) policy.
- periodic task sets with offsets scheduled according to the Fixed Priority Preemptive (FPP) policy [15].

The analysis for TDMA has mainly served to familiarize ourselves with the Prosa library. Schedulability analysis in presence of offsets is a non-trivial problem with a high computational complexity. In contrast to the traditional (offset oblivious) analysis, many scenarios must be tested and compared to identify which one represents the worst-case scenario. We have formalized and proved in Coq the basic analysis presented by Tindell [72]. This has allowed us to: (1) underline implicit assumptions made in Tindell's informal analysis; (2) ease the generalization of the verified analysis; (3) generate a certifier and an analyzer. We are investigating these two tools in terms of computational complexity and implementation effort, in order to provide a good solution to guarantee schedulability of industrial systems.

In parallel, we have worked on a Coq formalization of Typical Worst Case Analysis (TWCA). We aim to provide certified generic results for weakly-hard real-time systems in the form of (m, k) guarantees (a task may miss at most m deadlines out of k consecutive activations). So far, we have adapted the initial TWCA for arbitrary schedulers. The proof relies on a practical definition of the concept of busy window which amounts to being able to perform a local response time analysis. We provide such an instantiation for Fixed Priority Preemptive (FPP) schedulers as in the original paper. Future work includes making the state of the art TWCA suitable for formal proofs, exploring more complex systems (*e.g.*, bounded buffers) and providing instantiations of our results for other scheduling policies.

6.3. Language Based Fault-Tolerance

Participants: Pascal Fradet, Alain Girault, Gregor Goessler, Jean-Bernard Stefani, Martin Vassor.

6.3.1. Fault Ascription in Concurrent Systems

The failure of one component may entail a cascade of failures in other components; several components may also fail independently. In such cases, elucidating the exact scenario that led to the failure is a complex and tedious task that requires significant expertise.

The notion of causality *(did an event e cause an event e'?)* has been studied in many disciplines, including philosophy, logic, statistics, and law. The definitions of causality studied in these disciplines usually amount to variants of the counterfactual test "e is a cause of e' if both e and e' have occurred, and in a world that is as close as possible to the actual world but where e does not occur, e' does not occur either". In computer science, almost all definitions of logical causality — including the landmark definition of [54] and its derivatives — rely on a causal model that. However, this model may not be known, for instance in presence of black-box components. For such systems, we have been developing a framework for blaming that helps us establish the causal relationship between component failures and system failures, given an observed system execution trace. The analysis is based on a formalization of counterfactual reasoning [6].

In [16] we have discussed several shortcomings of existing approaches to counterfactual causality from the computer science perspective, and sketched lines of work to try and overcome these issues. In particular, research on counterfactual causality analysis has been marked, since its early days, by a succession of definitions of causality that are informally (in)validated against human intuition on mostly simple examples, see *e.g.*, [54], [53]. We call this approach TEGAR, *textbook example guided analysis refinement*. As pointed out in [48], it suffers from its dependence on the tiny number and incompleteness of examples in the literature, and from the lack of stability of the intuitive judgments against which the definitions are validated. We have argued that we need a formalization of counterfactual causality based on *first principles*, in the sense that causality definitions should not be driven by individual examples but constructed from a set of precisely specified requirements. Example of such requirements are robustness of causation under equivalence of models, and well-defined behavior under abstraction and refinement. To the best of our knowledge, none of the existing causality analysis techniques provides sufficient guarantees in this regard.

We are currently working on a revised version of our general semantic framework for fault ascription in [50] that satisfies a set of formally stated requirements, and on its instantiation to acyclic models of computation, in order to compare our approach with the standard definition of *actual causality* proposed by Halpern and Pearl.

6.3.2. Tradeoff exploration between energy consumption and execution time

We have continued our work on multi-criteria scheduling, in two directions. First, in the context of dynamic applications that are launched and terminated on an embedded homogeneous multi-core chip, under execution time and energy consumption constraints, we have proposed a two layer adaptive scheduling method [14]. In the first layer, each application (represented as a DAG of tasks) is scheduled statically on subsets of cores: 2 cores, 3 cores, 4 cores, and so on. For each size of these sets (2, 3, 4, ...), there may be only one topology or several topologies. For instance, for 2 or 3 cores there is only one topology (a "line"), while for 4 cores there are three distinct topologies ("line", "square", and "T shape"). Moreover, for each topology, we generate statically several schedules, each one subject to a different total energy consumption constraint, and consequently with a different Worst-Case Reaction Time (WCRT). Coping with the energy consumption constraints is achieved thanks to Dynamic Frequency and Voltage Scaling (DVFS). In the second layer, we use these pre-generated static schedules to reconfigure dynamically the applications running on the multi-core each time a new application is launched or an existing one is stopped. The goal of the second layer is to perform a dynamic global optimization of the configuration, such that each running application meets a pre-defined quality-of-service constraint (translated into an upper bound on its WCRT) and such that the total energy consumption be minimized. For this, we *(i)* allocate a sufficient number of cores to each active application, *(ii)* allocate the unassigned cores to the applications yielding the largest gain in energy, and *(iii)* choose for each application the best topology for its subset of cores (*i.e.*, better than the by default "line" topology). This is a joint work with Ismail Assayad (U. Casablanca, Morocco) who visits the team regularly.

Second, we have proposed the first of its kind multi-criteria scheduling heuristics for a DAG of tasks onto an homogeneous multi-core chip, optimizing the execution time, the reliability, the power consumption, and the temperature. Specifically, we have worked on the static scheduling minimizing the execution time of the application under the multiple constraints that the reliability, the power consumption, and the temperature remain below some given thresholds. There are multiple difficulties: *(i)* the reliability is not an invariant measure w.r.t. time, which makes it impossible to use backtrack-free scheduling algorithms such as list scheduling [28]; to overcome this, we adopt instead the Global System Failure Rate (GSFR) as a measure of the system's reliability, which is invariant with time [46]; *(ii)* keeping the power consumption under a given threshold requires to lower the voltage and frequency, but this has a negative impact both on the execution time and on the GSFR; keeping the GSFR below a given threshold requires to replicate the tasks on multiple cores, but this has a negative impact both on the execution time, on the power consumption, and on the temperature; *(iii)* keeping the temperature below a given threshold is even more difficult because the temperature continues to increase even after the activity stops, so each scheduling decision must be assessed not based on the current state of the chip (*i.e.*, the temperature of each core) but on the state of the chip at the end of the candidate task, and cooling slacks must be inserted. We have proposed a multi-criteria scheduling heuristics to address these challenges. It produces a static schedule of the given application graph and the given architecture description, such that the GSFR, power, and temperature thresholds are satisfied, and such that the execution time is minimized. We then combine our heuristic with a variant of the ε-constraint method [52] in order to produce, for a given application graph and a given architecture description, its entire Pareto front in the 4D space (exec. time, GSFR, power, temp.). This is a joint work with Athena Abdi and Hamid Zarandi from Amirkabir U., Iran, who have visited the team in 2016.

6.3.3. Concurrent flexible reversibility

Reversible concurrent models of computation provide natively what appears to be very fine-grained checkpoint and recovery capabilities. We have made this intuition clear by formally comparing a distributed algorithm for checkpointing and recovery based on causal information, and the distributed backtracking algorithm that lies at the heart of our reversible higher-order pi-calculus. We have shown that (a variant of) the reversible higher-order calculus with explicit rollback can faithfully encode a distributed causal checkpoint and recovery

algorithm. The reverse is also true but under precise conditions, which restrict the ability to rollback a computation to an identified checkpoint. This work has currently not been published.

TEA Project-Team

7. New Results

7.1. ADFG: Affine data-flow graphs scheduler synthesis

Participants: Loïc Besnard, Thierry Gautier, Alexandre Honorat, Jean-Pierre Talpin, Hai Nam Tran.

We consider with ADFG (Affine DataFlow Graph) the synthesis of periodic scheduling parameters for real-time systems modeled as ultimately cyclo-static dataflow (UCSDF) graphs [14]. This synthesis aims for a trade-off between throughput maximization and total buffer size minimization. The synthesizer inputs are: a UCSDF graph which describes tasks by their Worst Case Execution Time (WCET), and directed buffers connecting tasks by their data production and consumption rates; the number of processors in the target system and the real-time scheduling synthesis algorithm to be used. The outputs are the synthesized scheduling parameters: the tasks periods, offsets, processor bindings and priorities, and the buffers initial marking and maximum sizes.

ADFG was originally the implementation of Adnan Bouakaz's work [0]. However the tool had not been packaged yet to be easily installed and used. Moreover, code refactoring led to improve the theory, and to add new features. Firstly, more accurate bounds and Integer Linear Programming (ILP) formulations have been used. Besides, dataflow graphs do not need to be weakly connected for EDF policy on multiprocessor systems. The new implementation also avoids to use a fixed parameter for some multiprocessor partitioning algorithms, now an optional strategy enables to compute it. Finally implementation has been adapted to standard technologies to be more easily installed and used. As the synthesizer evolved a lot, new evaluations have been made. Moreover, many scheduled examples have been simulated with Cheddar [0], which provides pertinent metrics to analyze the scheduling efficiency.

ADFG is being extended to investigate and solve the scheduling problem of dataflow programs on many-core architectures. These architectures have distinctive traits requiring significant changes to classical multiprocessor scheduling theory. There is a high number of contention points introduced by novel memory architectures and new interconnect types such as Network-on-Chip. Two solutions are proposed and implemented in ADFG: contention-aware and contention-free scheduling synthesis. We either take into account the contention and synthesize a contention-aware schedule or find a schedule that results in no contention.

7.2. Formal Semantics of Behavior Specifications in the Architecture Analysis and Design Language Standard

Participants: Loïc Besnard, Thierry Gautier, Jean-Pierre Talpin.

The Architecture Analysis and Design Language (AADL) is a standard proposed by SAE to express architecture specifications and share knowledge between the different stakeholders about the system being designed. To support unambiguous reasoning, formal verification, high-fidelity simulation of architecture specifications in a model-based AADL design workflow, we have defined formal semantics for the behavior specification of the AADL. These semantics rely on the structure of automata present in the standard already, yet provide tagged, trace semantics framework to establish formal relations between (synchronous, timed, asynchronous) usages or interpretations of behavior [17]. We define the model of computation and communication of a behavior specification by the synchronous, timed or asynchronous traces of automata with variables. These constrained automata are derived from *polychronous automata* defined within the polychronous model of computation and communication [11].

[0] *Real-Time Scheduling of Dataflow Graphs.* A. Bouakaz. PhD Thesis, University of Rennes 1, 2013.
[0] *The Cheddar project: a GPL real-time scheduling analyzer.* http://beru.univ-brest.fr/~singhoff/cheddar/

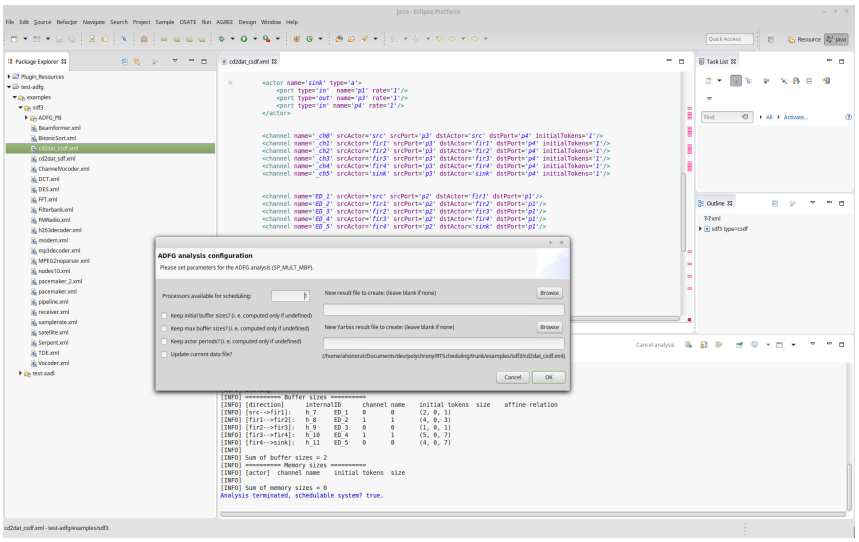

Figure 1. ADFG under Eclipse

States of a behavior annex transition system can be either observable from the outside (*initial*, *final* or *complete* states), that is states in which the execution of the component is paused or stopped and its outputs are available; or non observable execution states, that is internal states. We thus define two kinds of steps in the transition system: *small steps*, that is non-observable steps from or to an internal state; and *big steps*, that is observable steps from a *complete* state to another, through a number of small steps). The semantics of the AADL considers the observable states of the automaton. The set of states S_A of automaton A (used to interpret the behavior annex) thus only contains states corresponding to these observable states and the set of transitions T_A big-step transitions from an observable state to another (by opposition with small-step transitions from or to an execution state). The action language of the behavior annex defines actions performed during transitions. Actions associated with transitions are action blocks that are built from basic actions and a minimal set of control structures (sequences, sets, conditionals and loops). Typically, a behavior action sequence is represented by concatenating the transition systems of its elements; a behavior action set is represented by composing the transition systems of its elements.

The polychronous model of computation had been used previously as semantic model for systems described in the core AADL standard. This translation of AADL specifications into the polychronous model now takes into account the behavior specifications. The import of AADL behavior annexes (AADL-BA) to the polychronous model relies on polychronous automata and on small steps/big steps semantics. Small steps may be viewed as an implicit oversampling of the big steps. To express such implicit upsampling, a model of *Signal-thread* has been introduced in Polychrony (refer to Section “New trends and developments in Polychrony”). In that context, the translation of a behavior annex associated with an AADL thread consists mainly in the production of the corresponding Signal automaton, which is declared as a Signal-thread, and the definition of the environment required for this Signal-thread. In particular, the signal *complete-thread* is defined so that it will occur when the next state of the automaton is a *complete* state (the control will return to the scheduler): in other words, it specifies the end of a sequence of small steps.

A specific difficulty in the translation of AADL-BA is the translation of the action language, which is related to the general problem of the translation of a sequential language to a dataflow one. First, in AADL-BA actions, a given variable may be assigned several times in a sequence (for example, $x = a + b; x = x + a$). Thus an AADL-BA action has to be transformed into a SSA (static single assignment) form ($x_0 = a + b; x = x_0 + a$

in the previous example). Another possible problem is the translation of AADL-BA loop structures (for, while, do until). In our case, this is solved, again, by considering them as Signal-threads: the *dispatch-thread* event is defined by the upperbound of the clocks of the inputs of the loop and the *complete-thread* event defines the termination of the loop.

7.3. New trends and developments in Polychrony

Participants: Loïc Besnard, Thierry Gautier.

The synchronous modeling paradigm provides strong correctness guarantees for embedded system design while requiring minimal environmental assumptions. In most related frameworks, global execution correctness is achieved by ensuring the insensitivity of (logical) time in the program from (real) time in the environment. This property, called endochrony, can be statically checked, making it fast to ensure design correctness. Unfortunately, it is not preserved by composition, which makes it difficult to exploit with component-based design concepts in mind.

It has been shown that compositionality can be achieved by weakening the objective of endochrony: a weakly endochronous system is a deterministic system that can perform independent computations and communications in any order as long as this does not alter its global state. Moreover, the non-blocking composition of weakly endochronous processes is isochronous, which means that the synchronous and asynchronous compositions of weakly endochronous processes accept the same behaviors. Unfortunately, testing weak endochrony needs state-space exploration, which is very costly in the general case. Then, a particular case of weak endochrony, called polyendochrony, was defined, which allows static checking thanks to the existing clock calculus. The clock hierarchy of a polyendochronous system may have several trees, with synchronization relations between clocks placed in different trees, but the clock expressions of the clock system must be such that there is no clock expression (especially, no root clock expression) defined by symmetric difference: root clocks cannot refer to absence. In other words, the clock system must be in disjunctive form [9].

We have now implemented code generation for polyendochronous systems in Polychrony. This generation reuses techniques of distributed code generation, with rendez-vous management for synchronization constraints on clocks which are not placed in the same tree of clocks. For such a synchronization constraint $c_1 = c_2$, nodes *send* and *receive* are added in the graph, associated with clocks c_1 and c_2: for c_1, *send*(c_1) is followed by *receive*(c_2), followed itself by all the other nodes associated with clock c_1; and symmetrically for c_2. Then the subgraphs corresponding respectively to the trees where c_1 and c_2 are placed are separated, as if they were distributed on different processors. In this way, nodes *send* and *receive* become respectively outputs and inputs (both for c_1 and c_2) of the subgraphs. Finally, a communication library (MPI) is used for simulation. The following restriction is considered in the current implementation: the roots of the trees of c_1 and c_2 must be free variables.

We have also considered another extension related to clocks, again for making code generation possible for more programs than it was the case before. A characteristic of the Signal language is that it allows to specify programs which have internal accelerations with respect to their inputs and outputs. However, the constraint that implemented programs, for which code was generated, should be endochronous, restricted more or less these programs to have one single such acceleration (or clock upsampling). To abstract from this restriction, we have defined a model of so-called *Signal-thread*, that helps to confine such accelerations, and thus to generate code for programs with multiple clock upsampling. A Signal-thread is a Signal process with internal implicit upsampling; it has a *dispatch-thread* input event and a *complete-thread* output event; its outputs are delayed compared with its inputs. As the Signal-thread represents an upsampling, the *step* (see [1]) of the corresponding generated code is a loop. Such Signal-threads may be considered as a pragmatic way to implement *clock domains*.

7.4. Modular verification of cyber-physical systems using contract theory

Participants: Jean-Pierre Talpin, Benoit Boyer, David Mentre, Simon Lunel.

The primary goal of our project, in collaboration with Mitsubishi Electronics Research Centre Europe (MERCE), is to ensure correctness-by-design in realistic cyber-physical systems, i.e., systems that mix software and hardware in a physical environment, e.g., Mitsubishi factory automation lines or water-plant factory. To achieve that, we develop a verification methodology based on decomposition into components enhanced with contract reasoning.

The work of A. Platzer on Differential Dynamic Logic ($d\mathcal{L}$) holds our attention [0]. This a formalism built on the Dynamic Logic of V. Pratt augmented with the possibility of expressing Ordinary Differential Equations (ODEs), which are the usual way to model physical behaviors in physics. Combined with the ability of Dynamic Logic to specify and verify hybrid programs, $d\mathcal{L}$ is particularly fit model cyber-physical systems. The proof system associated with the logic is implemented into the theorem prover KeYmaera X. Aimed toward automatisation, it is a promising tool to spread formal methods into industry.

We have defined a syntactic parallel composition operator in $d\mathcal{L}$ which enjoys associativity and commutativity [15]. Commutativity allows to compose component in every possible order. Associativity is mandatory to modularly design a system; it allow to upgrade a system by adding new components. We have then characterized the conditions under which we can derive automatically a proof of the contract of our composition of two components, given the proof of the contract for each components. Theses theoretical results have been exemplified with an example of a cruise-controller entirely proved within the interactive theorem prover KeYmaera X.

The study of the cruise-controller example and of a water-tank system highlights some limitations of our approach. We can not handle retro-action and we have to compose in parallel components which have to be sequenced, e.g. a sensor and a computer. We have overcomed theses limitations by introducing a sequential composition operator which enjoys associativity and distributivity over the parallel composition operator. We believe it is a first step toward a composition algebra in $d\mathcal{L}$. This operator also satisfy the property that we can automatically derive a proof of the contract of our composition of two components, given the proof of the contract for each components, but under some relaxed conditions. We believe it is the first step toward a composition algebra.

Thanks to these results, a wide variety of systems are now possible to modularly design in $d\mathcal{L}$. To validate our approach, we are currently working on the implementation of our parallel composition operator as a tactic in KeYmaera X.

To challenge our ideas, we are working in the proof of a realistic cyber-physical system, a power-train system used in automotive. We plan to use it as a basis to test abstraction mechanisms to ultimately allow mix between top-down and bottom-up design.

7.5. Parametric verification of time synchronization protocols

Participants: Ocan Sankur, Jean-Pierre Talpin.

In the context of the associate-team COMPOSITE, we addressed the verification one of the apparently simplest services in any loosely-coupled distributed system : the time service. In many instances of such systems, traffic and power grids, banking and transaction networks, the accuracy and reliability of this service are critical.

In the instance of sensor networks, it is of particular interest to verify the robustness of such protocols to variations caused by the environment. Lake of power, varying temperatures, imperfect hardware, are sources of local drifts and jitters in time measurement that require self-calibration and fault-tolerance to reach distributed consensus. FTSP, the flooding time synchronization protocol, provides fault-tolerance and enables time synchronization.

In [16], we introduce an environment abstraction technique and an incremental model checking technique to prove that FTSP eventually elects a leader for any network topology and configuration (anonymized identifiers), up to a diameter $N = 7$ (with synchronous communications) and $N = 5$ (desynchronized communications), resulting in significant improvements over previous results.

[0] *Differential Dynamic Logic for Hybrid Systems*, André Platzer, http://symbolaris.com/logic/dL.html

7.6. Modular analysis and verification of system libraries

Participants: Jean-Joseph Marty, Jean-Pierre Talpin.

We are starting to develop a new perspective on the active topic of information flow control (IFC). We plan to adapt current investigations to tagged multi-core architecture, including software (virtual machines) and hardware (the Risc V processor) experiments and applications. All this work is based on the previous experience about verified Unikernel programming on low resources processors such as the Arduino (Marty's Master internship). We will define formally relations between processes and blocks of code inside a concurrent environment. This line of work will be investigated for both embedded IoT applications and cloud computing. By working with IFC at processor level and system level, we will enforce strong security foundation and focus on constraint solving analysed software.

ANTIQUE Project-Team

7. New Results

7.1. Memory Abstraction

7.1.1. Abstraction of arrays based on non contiguous partitions

Participants: Jiangchao Liu, Xavier Rival [correspondant].

In [9], we studied array abstractions.

Array partitioning analyses split arrays into contiguous partitions to infer properties of cell sets. Such analyses cannot group together non contiguous cells, even when they have similar properties. We proposed an abstract domain which utilizes semantic properties to split array cells into groups. Cells with similar properties will be packed into groups and abstracted together. Additionally, groups are not necessarily contiguous. This abstract domain allows to infer complex array invariants in a fully automatic way. Experiments on examples from the Minix 1.1 memory management demonstrated its effectiveness.

7.1.2. Semantic-Directed Clumping of Disjunctive Abstract States

Participants: Huisong Li, Francois Berenger, Bor-Yuh Evan Chang, Xavier Rival [correspondant].

In [16], we studied the semantic directed clumping of disjunctive abstract states.

To infer complex structural invariants, Shape analyses rely on expressive families of logical properties. Many such analyses manipulate abstract memory states that consist of separating conjunctions of basic predicates describing atomic blocks or summaries. Moreover, they use finite disjunctions of abstract memory states in order to account for dissimilar shapes. Disjunctions should be kept small for the sake of scalability, though precision often requires to keep additional case splits. In this context, deciding when and how to merge case splits and to replace them with summaries is critical both for the precision and for the efficiency. Existing techniques use sets of syntactic rules, which are tedious to design and prone to failure. In this paper, we design a semantic criterion to clump abstract states based on their silhouette which applies not only to the conservative union of disjuncts, but also to the weakening of separating conjunction of memory predicates into inductive summaries. Our approach allows to define union and widening operators that aim at preserving the case splits that are required for the analysis to succeed. We implement this approach in the MemCAD analyzer, and evaluate it on real-world C codes from existing libraries, including programs dealing with doubly linked lists, red-black trees and AVL-trees.

7.1.3. Relational Inductive Shape Abstraction

Participants: Hugo Illous, Matthieu Lemerre, Xavier Rival [correspondant].

In [13], we studied a relational inductive shape abstract domain.

Static analyses aim at inferring semantic properties of programs. While many analyses compute an over-approximation of reachable states, some analyses compute a description of the input-output relations of programs. In the case of numeric programs, several analyses have been proposed that utilize relational numerical abstract domains to describe relations. On the other hand, designing abstractions for relations over memory states and taking shapes into account is challenging. In this paper, we propose a set of novel logical connectives to describe such relations, which are inspired by separation logic. This logic can express that certain memory areas are unchanged, freshly allocated, or freed, or that only part of the memory was modified. Using these connectives, we build an abstract domain and design a static analysis that over-approximates relations over memory states containing inductive structures. We implement this analysis and report on the analysis of a basic library of list manipulating functions.

7.2. Static Analysis of JavaScript Code

7.2.1. Weakly Sensitive Analysis for Unbounded Iteration over JavaScript Objects

Participants: Yoonseok Ko, Xavier Rival [correspondant], Sukyoung Ryu.

In [14], we studied composite object abstraction for the analysis JavaScript.

JavaScript framework libraries like jQuery are widely use, but complicate program analyses. Indeed, they encode clean high-level constructions such as class inheritance via dynamic object copies and transformations that are harder to reason about. One common pattern used in them consists of loops that copy or transform part or all of the fields of an object. Such loops are challenging to analyze precisely, due to weak updates and as unrolling techniques do not always apply. In this work, we observe that precise field correspondence relations are required for client analyses (e.g., for call-graph construction), and propose abstractions of objects and program executions that allow to reason separately about the effect of distinct iterations without resorting to full unrolling. We formalize and implement an analysis based on this technique. We assess the performance and precision on the computation of call-graph information on examples from jQuery tutorials.

7.2.2. Revisiting recency abstraction for JavaScript: towards an intuitive, compositional, and efficient heap abstraction

Participants: Jihyeok Park, Xavier Rival [correspondant], Sukyoung Ryu.

In [18], we studied recency abstractions and their use for the analysis of JavaScript programs.

JavaScript is one of the most widely used programming languages. To understand the behaviors of JavaScript programs and to detect possible errors in them, researchers have developed several static analyzers based on the abstract interpretation framework. However, JavaScript provides various language features that are difficult to analyze statically and precisely such as dynamic addition and removal of object properties, first-class property names, and higher-order functions. To alleviate the problem, JavaScript static analyzers often use recency abstraction, which refines address abstraction by distinguishing recent objects from summaries of old objects. We observed that while recency abstraction enables more precise analysis results by allowing strong updates on recent objects, it is not monotone in the sense that it does not preserve the precision relationship between the underlying address abstraction techniques: for an address abstraction A and a more precise abstraction B, recency abstraction on B may not be more precise than recency abstraction on A. Such an unintuitive semantics of recency abstraction makes its composition with various analysis sensitivity techniques also unintuitive. In this paper, we propose a new singleton abstraction technique, which distinguishes singleton objects to allow strong updates on them without changing a given address abstraction. We formally define recency and singleton abstractions, and explain the unintuitive behaviors of recency abstraction. Our preliminary experiments show promising results for singleton abstraction.

7.3. Astrée and AstréeA

7.3.1. Finding All Potential Run-Time Errors and Data Races in Automotive Software

Participants: Antoine Miné, Laurent Mauborgne, Xavier Rival [correspondant], Jerome Feret, Patrick Cousot, Daniel Kästner, Stephan Wilhelm, Christian Ferdinand.

Safety-critical embedded software has to satisfy stringent quality requirements. All contemporary safety standards require evidence that no data races and no critical run-time errors occur, such as invalid pointer accesses, buffer overflows, or arithmetic overflows. Such errors can cause software crashes, invalidate separation mechanisms in mixed-criticality software, and are a frequent cause of errors in concurrent and multi-core applications. The static analyzer ASTRÉE has been extended to soundly and automatically analyze concurrent software. This novel extension employs a scalable abstraction which covers all possible thread interleavings, and reports all potential run-time errors, data races, deadlocks, and lock/unlock problems. When the analyzer does not report any alarm, the program is proven free from those classes of errors. Dedicated support for ARINC 653 and OSEK/AUTOSAR enables a fully automatic OS-aware analysis. In [15], we give an overview

of the key concepts of the concurrency analysis and report on experimental results obtained on concurrent automotive software. The experiments confirm that the novel analysis can be successfully applied to real automotive software projects.

7.4. Static analysis of signaling pathways

7.4.1. Formal and exact reduction for differential models of signaling pathways in rule-based languages

Participant: Ferdinanda Camporesi.

The behavior of a cell is driven by its capability to receive, propagate and communicate signals. Proteins can bind together on some binding sites. Post- translational modifications can reveal or hide some sites, so new interactions can be allowed or existing ones can be inhibited.

Due to the huge number of different bio-molecular complexes, we can no longer derive or integrate ODE models. A compact way to describe these systems is supplied by rule-based languages. However combinatorial complexity raises again when one attempt to describe formally the behavior of the models. This motivates the use of abstractions.

In this PhD thesis, we propose two methods to reduce the size of the models, that exploit respectively the presence of symmetries between sites and the lack of correlation between different parts of the system. The symmetries relates pairs of sites having the same capability of interactions. We show that this relation induces a bisimulation which can be used to reduce the size of the original model. The information flow analysis detects, for each site, which parts of the system influence its behavior. This allows us to cut the molecular species in smaller pieces and to write a new system. Moreover we show how this analysis can be tuned with respect to a context.

Both approaches can be combined. The analytical solution of the reduced model is the exact projection of the original one. The computation of the reduced model is performed at the level of rules, without the need of executing the original model.

7.4.2. Translating BNGL models into Kappa our experience

Participant: Kim Quyen Ly [correspondant].

So as to test the Kappa development tools on more examples, we translated the models provided with the BNGL distribution, into Kappa. In [20], we report about our experience. The translation was quite straightforward except for few interesting issues that we detail here. Firstly the use of static analysis has exposed some glitches in the modelling of some pathways in the models of the BNGL distribution. We explain how static analysis has helped us to detect, locate, and correct these flaws. Secondly, expanding BNGL rules using equivalent sites into rules with uniquely identified sites is not so easy when one wants to preserve faithfully the kinetics of interactions. We recall the semantics of BNGL for equivalent sites, and explain how to perform such translation.

7.4.3. Using alternated sums to express the occurrence number of extended patterns in site-graphs

Participants: Ferdinanda Camporesi, Jerome Feret [correspondant].

Site-graph rewriting languages as Kappa or BNGL supply a convenient way to describe models of signaling pathways. Unlike classical reaction networks, they emphasise on the biochemical structure of proteins. In [10], we use patterns to formalise properties about bio-molecular species. Intentionally, a pattern is a part of a species, but extensionally it denotes the multi-set of the species containing this pattern (with the multiplicity). Thus reasoning on patterns allows to handle symbolically arbitrarily big (if not infinite) multi-sets of species. This is a key point to design fast simulation algorithms or model reduction schemes. In this paper, we introduce the notion of extended patterns. Each extended pattern is made of a classical pattern and of a set of potential bonds between pairs of sites. Extended patterns have positive (when at least one of the potential bonds is

realised) and negative (when none is realised) instances. They are important to express the consumption and the production of patterns by the rules that may break cycles in bio-molecular species by side-effects. We show that the number of positive (resp. negative) instances of extended patterns may be expressed as alternated sums of the number of occurrences of classical patterns.

7.4.4. KaDE: a Tool to Compile Kappa Rules into (Reduced) ODE Models

Participants: Ferdinanda Camporesi, Jerome Feret [correspondant], Kim Quyen Ly.

In [11], we introduce the tool KaDe, that may be used to compile models written in Kappa in ODE. Kappa is a formal language that can be used to model systems of biochemical interactions among proteins. It offers several semantics to describe the behaviour of Kappa models at different levels of abstraction. Each Kappa model is a set of context-free rewrite rules. One way to understand the semantics of a Kappa model is to read its rules as an implicit description of a (potentially infinite) reaction network. KaDE is interpreting this definition to compile Kappa models into reaction networks (or equivalently into sets of ordinary differential equations). KaDE uses a static analysis that identifies pairs of sites that are indistinguishable from the rules point of view, to infer backward and forward bisimulations, hence reducing the size of the underlying reaction networks without having to generate them explicitly. In [11], we describe the main current functionalities of KaDE and we give some benchmarks on case studies. A complete tutorial and more complete benchmarks may be found at the following url: http://www.di.ens.fr/~feret/CMSB2017-tool-paper/.

CELTIQUE Project-Team

4. New Results

4.1. Higher-Order Process Calculi

Participants: Sergueï Lenglet, Alan Schmitt.

Sergueï Lenglet and Alan Schmitt, in collaboration with researchers at Wrocław university, designed a fully abstract encoding of the λ-calculus into HOcore, a minimal higher-order process calculus. This work has been published at LICS [37]. In parallel, Lenglet and Schmitt have formalized HOπ in Coq and showed that its bisimilarity is compatible using Howe's method. This work has been accepted for publication at CPP 2018 [30].

4.2. Certified Semantics and Analyses for JavaScript

Participants: Gurvan Cabon, Alan Schmitt.

Alan Schmitt has continued his collaboration with Arthur Charguéraud (Inria Nancy) and Thomas Wood (Imperial College London) to develop JSExplain, an interpreter for JavaScript that is as close as possible to the specification. The tool is publicly available at https://github.com/jscert/jsexplain and is being extended to cover the current version of the standard.

In parallel, Gurvan Cabon and Alan Schmitt have developed a framework to automatically derive an information-flow tracking semantics from a pretty-big-step semantics. This work has been published [34] and is being formalized in Coq.

4.3. Certified Concurrent Garbage Collector

Participants: Yannick Zakowski, David Cachera, Delphine Demange, David Pichardie.

Concurrent garbage collection algorithms are an emblematic challenge in the area of concurrent program verification. We addressed this problem by proposing a mechanized proof methodology based on the popular Rely-Guarantee (RG) proof technique. We designed a specific compiler intermediate representation (IR) with strong type guarantees, dedicated support for abstract concurrent data structures, and high-level iterators on runtime internals (objects, roots, fields, thread identifiers...). In addition, we defined an RG program logic supporting an incremental proof methodology where annotations and invariants can be progressively enriched. We have formalized the IR, the proof system, and proved the soundness of the methodology in the Coq proof assistant. Equipped with this IR, we have proved the correctness of a fully concurrent garbage collector where mutators never have to wait for the collector. This work has been published in [32].

In this work, reasoning simultaneously about the garbage collection algorithm and the concrete implementation of the concurrent data-structures it uses would have entailed an undesired and unnecessary complexity. The above proof is therefore conducted with respect to abstract operations which execute atomically. In practice, however, concurrent data-structures uses fine-grained concurrency, for performance reasons. One must therefore prove an observational refinement between the abstract concurrent data-structures and their fined-grained, "linearisable" implementation. To adress this issue, we introduce a methodology inspired by the work of Vafeiadis, and provide the approach with solid semantic foundations. Assuming that fine-grained implementations are proved correct with respect to an RG specification encompassing linearization conditions, we prove, once and for all, that this entails a semantic refinement of their abstraction. This methodology is instantiated to prove correct the main data-structure used in our garbage collector. This work has been published in [33].

4.4. Static analysis of functional programs using tree automata and term rewriting

Participants: Thomas Genet, Thomas Jensen, Timothée Haudebourg.

We develop a specific theory and the related tools for analyzing programs whose semantics is defined using term rewriting systems. The analysis principle is based on regular approximations of infinite sets of terms reachable by rewriting. Regular tree languages are (possibly) infinite languages which can be finitely represented using tree automata. To over-approximate sets of reachable terms, the tools we develop use the Tree Automata Completion (TAC) algorithm to compute a tree automaton recognizing a superset of all reachable terms. This over-approximation is then used to prove properties on the program by showing that some "bad" terms, encoding dangerous or problematic configurations, are not in the superset and thus not reachable. This is a specific form of, so-called, Regular Tree Model Checking. We have already shown that tree automata completion can safely over-approximate the image of any first-order complete and terminating functional program. We have extended this result to the case of higher-order functional programs [40] and obtained very encouraging experimental results http://people.irisa.fr/Thomas.Genet/timbuk/funExperiments/. Besides, we have shown that completion was abble to take the evaluation strategy of the program into account [19]. The next step is to show the completeness of the approach, i.e., that any regular approximation of the image of a function can be found using completion. We already made progress in this direction [39].

4.5. C Semantics and Certified Compilation

Participants: Frédéric Besson, Sandrine Blazy.

The CompCert C compiler provides the formal guarantee that the observable behaviour of the compiled code improves on the observable behaviour of the source code. A first limitation of this guarantee is that if the source code goes wrong, i.e. does not have a well-defined behaviour, any compiled code is compliant. Another limitation is that CompCert 's notion of observable behaviour is restricted to IO events.

Over the past years, we have refined the semantics underlying CompCert so that (unlike CompCert but like Gcc) the binary representation of pointers can be manipulated much like integers and such that memory is a finite resource. We have now a formally verified C compiler, CompCertS, which is essentially the CompCert compiler, albeit with a stronger formal guarantee. The semantics preservation theorem applies to a wider class of existing C programs and, therefore, their compiled version benefits from the formal guarantee of CompCertS. CompCertS preserves not only the observable behaviour of programs but also ensures that the memory consumption is preserved by the compiler. As a result, we have the formal guarantee that the compiled code requires no more memory than the source code. This ensures that the absence of stack-overflows is preserved by compilation.

The whole proof of CompCertS represents a significant proof-effort. Details about the formal definition of the semantics and the proof of compiler passes can be found in the following publications [17], [25]

4.6. Constant-time verification by compilation and static analysis

Participants: Sandrine Blazy, David Pichardie, Alix Trieu.

To protect their implementations, cryptographers follow a very strict programming discipline called constant-time programming. They avoid branchings controlled by secret data as an attacker could use timing attacks, which are a broad class of side-channel attacks that measure different execution times of a program in order to infer some of its secret values. Several real-world secure C libraries such as NaCl, mbedTLS, or Open Quantum Safe, follow this discipline. We propose an advanced static analysis, based on state-of-the-art techniques from abstract interpretation, to report time leakage during programming. To that purpose, we analyze source C programs and use full context-sensitive and arithmetic-aware alias analyses to track the tainted flows. We give semantic evidences of the correctness of our approach on a core language. We also present a prototype implementation for C programs that is based on the CompCert compiler toolchain and its companion Verasco static analyzer. We present verification results on various real-world constant-time programs and report on a successful verification of a challenging SHA-256 implementation that was out of scope of previous tool-assisted approaches. This work has been published at ESORICS'17 [27].

The previous technique is well-adapted to verify the constant-time discipline at source level and give feedback to programmers, but the final security property must be established on the executable form of the program. In a joint work with IMDEA Software (Gilles Barthe and Vincent Laporte), we propose an automated methodology for validating on low-level intermediate representations the results of a source-level static analysis. Our methodology relies on two main ingredients: a relative-safety checker, an instance of a relational verifier which proves that a program is *safer* than another, and a transformation of programs into defensive form which verifies the analysis results at runtime. We prove the soundness of the methodology, and provide a formally verified instantiation based on the Verasco verified C static analyzer and the CompCert verified C compiler. This work has been published at CSF'17 [24].

CONVECS Project-Team

6. New Results

6.1. New Formal Languages and their Implementations

LNT is a next generation formal description language for asynchronous concurrent systems, which attempts to combine the best features of imperative programming languages and value-passing process algebras. LNT is increasingly used by CONVECS for industrial case studies and applications (see § 6.5) and serves also in university courses on concurrency, in particular at ENSIMAG (Grenoble) and at Saarland University.

6.1.1. Translation from LNT to LOTOS

Participants: Hubert Garavel, Frédéric Lang, Wendelin Serwe.

The move towards "safer" LNT exceptions initiated in 2016 has been completed in 2017: the two concepts of gates and exceptions have been unified in both LNT processes and LNT functions. The static semantics of LNT no longer requires that variables and exceptions share the same name space.

LNT now permits simple loops (of the form "loop ... end loop", without loop label nor "while" condition) in LNT functions, as well as in LNT processes.

The pragma names "comparedby", "external", "implementedby", "iteratedby", "printedby", and "representedby" are no longer reserved LNT keywords, meaning that it is now permitted to declare LNT identifiers having these names. Two new type pragmas "!card" and "!bits" have been added to specify the maximum number of values and the number of bits to be used when storing the values of a given type in "hash-consing" tables.

The LPP preprocessor and the LNT2LOTOS translator, which implement the LNT language, have been enhanced in many ways. In addition to 9 bug fixes, the following enhancements have been made:

- LPP now implements LNT character strings more concisely.
- LPP automatically adds the ".lnt" extension to input and output files if this extension is missing.
- The algorithm that computes which LNT gates are used in each function or process has been made more precise, and LNT2LOTOS now warns about gates that are declared but never used.
- LNT2LOTOS performs stricter compile-time checks that produce dedicated error messages, rather than generating invalid LOTOS code that was subsequently rejected by CAESAR and/or CAESAR.ADT. Also, several error messages displayed by LNT2LOTOS during its static-analysis phases have been enhanced.
- The translation from LNT functions to LOTOS operations has been significantly improved by eliminating unreachable or redundant LOTOS equations, removing unused auxiliary LOTOS operations, simplifying the premises of certain LOTOS equations, factorizing identical assignments in "if-then-else" instructions, and optimizing long sequences of assignments intertwined with assertions. Thus, LNT2LOTOS is now faster, uses less memory, generates more compact LOTOS code, and can compile larger LNT specifications that could not be handled before.

The LNT2LOTOS Reference Manual, which contains the definition of the LNT language, has been revised, enriched, and simplified in many ways. A paper presenting the historical background and motivation behind the definition of LNT was published in an international conference [18].

6.1.2. NUPN

Participant: Hubert Garavel.

Nested-Unit Petri Nets (NUPNs) is an upward-compatible extension of P/T nets, which are enriched with structural information on their concurrent structure. Such additional information can easily be produced when NUPNs are generated from higher-level specifications (e.g., process calculi); quite often, such information allows logarithmic reductions in the number of bits required to represent states, thus enabling verification tools to perform better. The principles of NUPNs are exposed in [29] and its PNML representation is described here [0].

In 2017, we studied an abstraction called *place fusion*, which takes advantage of the compositional, hierarchical structure of NUPNs. We formulated key theorems stating which properties are preserved or not under this abstraction. On the practical side, our collection of NUPN models grew to more than 8 000 benchmarks. Statistical studies were done on this collection to estimate the compression factor permitted by the NUPN model. A journal article providing an overview of NUPNs was written.

The NUPN model has been adopted by the Model Checking Contest and implemented in ten different tools developed in four countries. In 2017, the NUPN model was also adopted for the parallel problems of the RERS'2017 (*Rigorous Examination of Reactive Systems*) challenge [0].

6.1.3. Analysis of BPMN via Translation to LNT

Participants: Ajay Muroor Nadumane, Gwen Salaün.

Business process modeling is an important concern in companies and organizations. Formal analysis techniques are crucial to detect semantic issues in the corresponding models, or to help with their refactoring and evolution. However, business process development frameworks often fall short when it comes to go beyond simulation or syntactic checking of the models. To ensure a more robust development of business processes, we developed the VBPMN verification framework. It features several techniques for the automatic analysis of business processes modeled using BPMN, the de facto standard for business process modeling.

The business processes, described using a Web application compliant with BPMN 2.0, are transformed into an intermediate format called PIF (*Process Intermediate Format*). Then, from the PIF descriptions, models in LNT and model-specific verification scripts in SVL are generated. In the end, CADP is used to check either for functional properties of a given business process, or for the correctness of the evolution of a business process into another one. This latter kind of verification supported by VBPMN is particularly helpful in order to improve a process w.r.t. certain optimization criteria. A paper presenting these results was published in an international conference [16].

6.1.4. Translation of Term Rewrite Systems

Participants: Hubert Garavel, Lina Marsso.

We pursued the development undertaken in 2015 of a software platform for systematically comparing the performance of rewrite engines and pattern-matching implementations in algebraic specification and functional programming languages. Our platform reuses the benchmarks of the three Rewrite Engine Competitions (2006, 2009, and 2010). Such benchmarks are term-rewrite systems expressed in a simple formalism named REC, for which we developed automated translators that convert REC benchmarks into many languages, among which AProVE, Clean, Haskell, LNT, LOTOS, Maude, mCRL, MLTON, OCAML, Opal, Rascal, Scala, SML-NJ, Stratego/XT, and Tom.

In 2017, we revised and enhanced the largest REC benchmark, the MAA (*Message Authenticator Algorithm*), a Message Authentication Code used for financial transactions (ISO 8731-2) between 1987 and 2002. This model (13 sorts, 18 constructors, 644 non-constructors, and 684 rewrite rules) was proven to be confluent, and terminating. Implementations in thirteen different languages have been automatically derived from this model and used to validate 200 official test vectors for the MAA. These results led to a publication in an international conference [14].

[0] http://mcc.lip6.fr/nupn.php
[0] http://www.rers-challenge.org

We also corrected and/or enhanced several of the existing REC translators (e.g., Clean) and added support of CafeOBJ and compiled OCAML. A scientific paper on this study has been prepared.

6.1.5. Other Language Developments

Participants: Hubert Garavel, Frédéric Lang, Radu Mateescu, Wendelin Serwe.

The ability to compile and verify formal specifications with complex, user-defined operations and data structures is a key feature of the CADP toolbox since its very origins.

In 2017, we brought various enhancements to several compilers handling formal specification languages (LOTOS, MCL, XTL, and GRL):

- A buffer overflow and two out-of-bound array accesses have been corrected in both CAESAR and CAESAR.ADT. Two memory allocation bugs have also been corrected in CAESAR.ADT. The latter tool now generates C code that gives better diagnostic when the evaluation of a constant fails at run time (e.g., when it triggers an exception signal, or exhausts the stack or heap memory).
- In addition to two bug fixes, the warning and error messages displayed by MCL_EXPAND and XTL_EXPAND have been made more precise and stringent. The XTL libraries "walk" and "walk_nice" have been modified not to trigger the extra warnings recently introduced.
- The GRL2LNT translator takes as input a formal description in GRL of a GALS system and generates an equivalent LNT specification. A new version 1.1 of GRL2LNT has been released, which corrects a bug concerning the LNT code generated by the "`-merge`" option.

H. Garavel pursued the study of the most suitable axiomatization of signed integers undertaken in 2016. He reviewed a tenth of such Peano-like axiom systems, which he classified and evaluated according to complexity and efficiency criteria. These results have been published in an international conference [13].

6.2. Parallel and Distributed Verification

6.2.1. Distributed State Space Manipulation

Participants: Hubert Garavel, Wendelin Serwe.

For distributed verification, CADP provides the PBG format, which implements the theoretical concept of *Partitioned LTS* [34] and provides a unified access to an LTS distributed over a set of remote machines.

In 2017, many changes have been done to simplify the code of the CAESAR_NETWORK_1 communication library, which is the backbone of the distributed verification tools of CADP, as well as the code of other tools such as BCG_MIN, but most of these changes are not directly observable by end users. In addition to two bug fixes in CAESAR_NETWORK_1 and two other bug fixes in the BES_SOLVE tool, the error messages displayed by the various tools and the statistical information produced by the "`-stat`" option of BES_SOLVE have been made more concise and more informative.

6.2.2. Debugging of Concurrent Systems

Participants: Gianluca Barbon, Gwen Salaün.

Model checking is an established technique for automatically verifying that a model satisfies a given temporal property. When the model violates the property, the model checker returns a counterexample, which is a sequence of actions leading to a state where the property is not satisfied. Understanding this counterexample for debugging the specification is a complicated task for several reasons: (i) the counterexample can contain hundreds of actions, (ii) the debugging task is mostly achieved manually, and (iii) the counterexample does not explicitly highlight the source of the bug that is hidden in the model.

We proposed an approach that improves the usability of model checking by simplifying the comprehension of counterexamples. Our solution aims at keeping only actions in counterexamples that are relevant for debugging purposes. To do so, we first extract in the model all the counterexamples. Second, we define an analysis algorithm that identifies actions that make the model skip from incorrect to correct behaviours, making these actions relevant from a debugging perspective. Our approach is fully automated by a tool we implemented and applied on real-world case studies from various application areas for evaluation purposes. This work led to a publication in an international conference [11].

In 2017, we focused on extending our approach following three directions: (a) we introduced new notions to identify new types of relevant actions; (b) we developed a set of heuristics to extract these actions from counterexamples; (c) we proposed an alternative approach to focus on a broader range of properties (i.e., liveness properties). These new extensions have been integrated into our tool. A paper was submitted to an international journal.

6.3. Timed, Probabilistic, and Stochastic Extensions

6.3.1. Tools for Probabilistic and Stochastic Systems

Participants: Hubert Garavel, Jean-Philippe Gros, Frédéric Lang, Julie Parreaux, Wendelin Serwe.

Formal models and tools dealing with quantititative aspects (such as time, probabilities, and other continuous physical quantities) have become unavoidable for a proper study and computer-aided verification of functional and non-functional properties of cyberphysical systems. The wealth of such formal models is sometimes referred to as a quantitative "zoo" [39].

The CADP toolbox already implements some of these probabilistic/stochastic models, namely DTMCs and CTMCs (*Discrete-Time* and *Continuous-Time Markov Chains*), and IMCs (*Interactive Markov Chains*) [41]. Our long-term goal is to increase the capability and flexibility of the CADP tools, so as to support other quantitative models more easily.

In 2017, we undertook a systematic review of the existing theoretical models and built a comprehensive list of more than 70 software tools implementing these models [37]. The results of this study have been made widely available as a Web catalog [0].

In parallel, we also undertook a systematic review [50] of the benchmarks made available for these tools. We downloaded more than 21 000 files from the web and developed triage scripts to analyze these files and classify them automatically, separating various kinds of automata-based models (e.g., Markov chains, Markov automata, hybrid automata, etc.) from temporal-logic formulas. One finding of this "big data" study is the present lack of diversity, as four tools (PRISM, MRMC, STORM, and SiSAT) provide nearly 60% of the models.

To address this issue, we started investigating the probabilistic and stochastic models of complex industrial systems produced by former PhD students of the VASY and CONVECS teams. We analyzed these models (written in BCG, EXP, LOTOS, LNT, SVL, and/or Makefiles) to separate functional aspects from performance ones, leading to a collection of DTMCs, CTMCs, IMCs, and IPCs (*Interactive Probabilistic Chains*). We updated these models to ensure compatibility with the latest versions of CADP and C compilers, and we started enhancing EXP.OPEN with new features that simplify the parallel composition of IPCs (see § 6.4.1).

6.3.2. On-the-fly Model Checking for Extended Regular Probabilistic Operators

Participant: Radu Mateescu.

[0] http://cadp.inria.fr/resources/zoo

Specifying and verifying quantitative properties of concurrent systems requires expressive and user-friendly property languages combining temporal, data-handling, and quantitative aspects. In collaboration with José Ignacio Requeno (Univ. Zaragoza, Spain), we undertook the quantitative analysis of concurrent systems modeled as PTSs (*Probabilistic Transition Systems*), whose actions contain data values and probabilities. We proposed a new regular probabilistic operator that extends naturally the Until operators of PCTL (*Probabilistic Computation Tree Logic*) [38], by specifying the probability measure of a path characterized by a generalized regular formula involving arbitrary computations on data values. We integrated the regular probabilistic operator into MCL, we devised an associated on-the-fly model checking method based on a combined local resolution of linear and Boolean equation systems, and we implemented the method in a prototype extension of the EVALUATOR model checker.

In 2017, we continued experimenting the extended model checker on further examples of protocols (Bounded Retransmission Protocol, randomized philosophers, self-stabilization) and observed that it exhibits a performance comparable with the explicit-state algorithms of the PRISM model checker [0]. A paper was submitted to an international journal.

6.4. Component-Based Architectures for On-the-Fly Verification

6.4.1. Compositional Verification

Participants: Hubert Garavel, Frédéric Lang.

The CADP toolbox contains various tools dedicated to compositional verification, among which EXP.OPEN, BCG_MIN, BCG_CMP, and SVL play a central role. EXP.OPEN explores on the fly the graph corresponding to a network of communicating automata (represented as a set of BCG files). BCG_MIN and BCG_CMP respectively minimize and compare behavior graphs modulo strong or branching bisimulation and their stochastic extensions. SVL (*Script Verification Language*) is both a high-level language for expressing complex verification scenarios and a compiler dedicated to this language.

In 2017, two bugs have been solved in SVL and one bug has been solved in EXP.OPEN. Several improvements have been brought to both tools. In particular:

- EXP.OPEN now has two new options "`-prob`" and "`-rate`" for handling probabilistic and stochastic transitions, respectively; without these options, probabilistic and stochastic transitions are considered as ordinary transitions (this enables EXP.OPEN to be used for implementing alternative semantics, such as *Interactive Probabilistic Chains* [27] where probabilistic transitions are synchronized using a global clock). Consequently, the former "`-ratebranching`" option has been replaced by "`-rate -branching`".

  Also, error messages about synchronization vectors have been made more precise and EXP.OPEN performs tighter checks about labels containing only blanks and unexpected synchronization of probabilistic or stochastic transitions. Two bugs have been fixed in EXP.OPEN and style files have been added to bring support for the EXP format by mainstream text editors.

- A new option "`-v`" has been added to set SVL variables from the command line (similar to "awk" or "make"). Debugging SVL scripts has been made easier: the "`-debug`" option of SVL now stops the execution as soon as a shell command (e.g., a CADP tool or a Unix command) terminates with a non-zero exit status, so that problems are detected as soon as they occur.

  Also, SVL now performs tighter semantic checks, making sure that all partial-order reduction options passed to EXP.OPEN (namely, options explicitly set by the user and options automatically computed by SVL from the context of the EXP composition expression) are not contradictory.

6.4.2. On-the-Fly Test Generation

Participants: Hubert Garavel, Lina Marsso, Radu Mateescu, Wendelin Serwe.

[0] http://www.prismmodelchecker.org/

The CADP toolbox provides support for conformance test case generation by means of the TGV tool. Given a formal specification of a system and a test purpose described as an input-output LTS (IOLTS), TGV automatically generates test cases, which assess using black box testing techniques the conformance of a system under test w.r.t. the formal specification. A test purpose describes the goal states to be reached by the test and enables one to indicate parts of the specification that should be ignored during the testing process. TGV does not generate test cases completely on the fly (i.e., *online*), because it first generates the complete test graph (CTG) and then traverses it backwards to produce controllable test cases.

In 2017, we carried out the following activities:

- We developed the prototype tool TESTOR to extract test cases completely on the fly. Compared to TGV, the new tool TESTOR presents several advantages: (i) it has a more modular architecture, based on generic graph transformation components taken from the OPEN/CAESAR libraries (τ-compression, τ-confluence, τ-closure, determinization, resolution of Boolean equation systems); (ii) it is capable of extracting a test case completely on the fly, by exploiting the diagnostic generation features of the Boolean equation system resolution algorithms; (iii) it enables a more flexible expression of test purposes, taking advantage of the multiway rendezvous, a primitive to express communication and synchronization among a set of distributed processes [15]. We evaluated TESTOR on three published case studies and more than 10 000 examples taken from the non-regression test suites of CADP. A paper describing this work was accepted for publication in an international conference.
- We also revised TGV, which is now by default much less verbose and only displays the most important information, but the former behaviour can still be retained using option "`-verbose`". A new option "`-monitor`" allows to follow in real time how the test case generation progresses. Many warning and error messages have been enhanced, various bugs (especially buffer overflows) have been fixed, and memory allocation results are now strictly controlled.

6.4.3. Other Component Developments

Participants: Lian Apostol, Soren Enevoldsen, Hubert Garavel, Frédéric Lang, Radu Mateescu, Wendelin Serwe.

The CAESAR_STANDARD library was enriched with the new CAESAR_TYPE_FORMAT type and its associated primitives, and with two new functions CAESAR_SET_SIGNALS() and CAESAR_RESET_SIGNALS() for handling POSIX signals (including SIGSEGV, i.e., segmentation violation). The CAESAR_GRAPH interface, which remained stable for two decades, has been modified: its two functions CAESAR_FORMAT_STATE() and CAESAR_FORMAT_LABEL() became more powerful, while its two functions CAESAR_MAX_FORMAT_STATE() and CAESAR_MAX_FORMAT_LABEL() have been removed from the interface. The same changes apply as well to all the other similar functions of the OPEN/CAESAR libraries. All the OPEN/CAESAR compilers, application tools, and demo examples have been modified to reflect these changes.

Sustained effort has been made to ensure that CADP works properly on mainstream computing platforms. In particular, the RFL and TST scripts and the documentation have been continuously updated. Changes were brought to CADP to cope with recent C compilers (such as GCC 6 and Clang) and to work around problems with the "indent" command available on Solaris and macOS/Xcode. On Linux, CADP was ported to the latest versions of Centos, Debian 9, and Ubuntu 17.04. The support for the various desktop environments (Gnome, KDE, Mate, etc.) available in Linux distributions has improved. On macOS, support of obsolete versions (from Mac OS X 10.6 "Snow Leopard" to OS X 10.9 "Mavericks" included) was withdrawn and support of macOS 10.13 "High Sierra" was added. Preliminary steps have been made to prepare a 64-bit version of CADP on macOS. On Windows, support of obsolete versions (Windows XP and Vista) was dropped. CADP was also adapted to follow the changes in the Cygwin software regarding pipe management. Many changes were made to CADP so as to support the case where Cygwin is not installed in "`C:/`" but in a different folder. Finally, preliminary steps have been made towards a 64-bit version of CADP for Windows.

In collaboration with Soren Enevoldsen (Aalborg University, Denmark), we studied the applicability of CADP tools for analyzing concurrent systems described using weighted CCS (WCCS) [43], an extension of CCS with an action prefix operator carrying a weight represented as a natural number. We developed a prototype OPEN/CAESAR-compliant compiler for WCCS, which enables to produce, in conjunction with the GENERATOR tool of CADP, the corresponding LTS model in which transitions are labeled with actions and weights. For specifying temporal properties of WCCS systems, we developed a prototype MCL library defining the operators of weighted CTL (WCTL) [43] using MCL fixed point operators parameterized by natural numbers. This library, used in conjunction with the EVALUATOR tool, provides an on-the-fly model checker for WCTL equipped with diagnostic capabilities (counterexamples and witnesses).

6.5. Real-Life Applications and Case Studies

6.5.1. Autonomous Resilience of Distributed IoT Applications in a Fog Environment

Participants: Umar Ozeer, Gwen Salaün.

The first year of the PhD thesis started with a state of the art on the resilience mechanisms, broadly in distributed systems and then more specifically in distributed IoT (*Internet of Things*) applications. This resulted, firstly in defining the scope of the thesis and, secondly, in identifying the steps to manage failures, namely state saving, failure detection, fault isolation, and repairing in a consistent state.

A study of the mechanisms for saving the execution state of processes in distributed systems was done. This enabled us to identify the specificities of our environment and to adapt existing snapshot and message logging mechanisms to fit the context of state saving and manipulation in distributed IoT applications in view of repairing failures and re-establishing consistency. We devised a first failure management protocol, which is being tested on an instance of an IoT application test bed at Orange Labs. Next steps include formally verifying the failure management protocol, as well as carrying out further tests on larger scaled applications for the purpose of performance evaluation.

6.5.2. Activity Detection in a Smart Home

Participants: Waqas Imtiaz, Frédéric Lang, Radu Mateescu, Wendelin Serwe.

Ambient intelligence is an active research field, whose aim is to design and analyze smart environments that are capable of automated interaction with users and the physical world, through sensors, actuators, displays, and computational elements, embedded in everyday objects, and connected through a network. In the Grenoble area, the Equipex Amiqual4Home [0] provides among others access to a Smart Home, which is a fully functional two-stage 90 meters square flat with 4 rooms including an open to kitchen living room, a bedroom, a bathroom and a small office. All the rooms are equipped with cameras, microphones, sensors and actuators to remote control various appliances like rollershutter, lights or multimedia devices. The software architecture of the Smart Home is based on the open source home automation software OpenHAB [0]. It allows a complete control of the flat devices with a single system, despite the various protocols used. Using the rule engine, it also enables the definition of rules expressing how the Smart Home should react to physical (human action, sensors, etc.) or external (weather prediction service, calendar, etc.) events. A difficult question is how to make sure that smart environments are programmed correctly, and will not lead to unexpected or even harmful behaviour.

Smart environments are concurrent and asynchronous by nature. To address the question above, we started, in collaboration with Nicolas Bonnefond (PERVASIVE INTERACTION team and Amiqual4Home), to study how existing tools for the formal design and verification of concurrent asynchronous systems present in the CADP toolbox can be used to verify a smart environment. Firstly, we proposed a translation from OpenHAB rules into a formal LNT model on which properties can be verified [42]. Secondly, in collaboration with Paula Lago and Claudia Roncancio (SIGMA team of LIG), we exploited the dataset ContextAct@A4H of daily living activities collected and annotated within Amiqual4Home for the purpose of activity recognition. Each activity was described as an MCL temporal logic formula that is checked repeatedly on the log of sensor

[0] http://amiqual4home.inria.fr
[0] http://www.openhab.org

measurements until all occurrences of the activity have been found. This approach has the ability to recognize the start and end points of activities (thus not requiring to segment sensor data) and also expresses the temporal order of events, thus palliating a limitation of existing ontology based activity recognition techniques. This led to a publication in an international conference [17].

6.5.3. Other Case Studies

Participants: Hubert Garavel, Frédéric Lang, Lina Marsso, Wendelin Serwe.

The demo examples of CADP, which have been progressively accumulated since the origins of the toolbox, are a showcase for the multiple capabilities of CADP, as well as a test bed to assess the new features of the toolbox. In 2017, the effort to maintain and enhance these demos has been pursued. The demo 05 (Airplane-ground communication protocol) has been modified to use the new syntax of exceptions in the LNT language. The LOTOS and LNT specifications of demo 12 (Message Authenticator Algorithm) have been entirely revised, based upon the fine knowledge acquired by modelling this cryptographic function as a term rewrite system [14]. The LNT specification has also been extended to incorporate the test vectors given in the International Standards ISO 8730 and 8731-2. The resulting specification, which was initially too large to be compiled, is now successfully handled after the enhancements brought to the LNT2LOTOS translator. Demo 19 (Production Cell) has been simplified and is now fully documented in a publication [15].

In the framework of the SECURIOT-2 project (see § 8.2.2.1), a Memory Protection Unit has been formally specified in LNT and verified at Tiempo using CADP. A paper has been submitted to an international conference.

DEDUCTEAM Project-Team

6. New Results

6.1. $\lambda\Pi$-calculus modulo theory

G. Dowek has given a semantic criterion for the termination of the $\lambda\Pi$-calculus modulo theory. This result has been published in [23].

A. Assaf, G. Dowek, J.-P. Jouannaud and J. Liu have given a confluence criterion for untyped higher-order rewrite systems, and demonstrated some applications to the $\lambda\Pi$-calculus modulo theory.

G. Dowek has given an invited talk at PxTP where he has presented a state of the art of the production of system-independent proof libraries. This paper has been published in the proceedings of PxTP [12].

6.2. Dedukti

During his internship [22], A. Defourné extended F. Blanqui's prototype of proof assistant based on Dedukti by developing a tactic for calling external provers through Why3 [28]. He also started to study a simple rewriting tactic.

During his internship, R. Bocquet studied unification in the $\lambda\Pi$-calculus modulo rewriting, and started to implement a prototype.

During his internship [24], G. Genestier studied the possibility to use the Size-Change Principle [34] in order to prove termination in the $\lambda\Pi$-calculus modulo rewriting. This work led to an adaptation of the criterion developped in his thesis by Wahlstedt [40] to a calculus containing dependant types. He also implemented a prototype of a weak version of the criterion.

During the first three months of his postdoc, R. Lepigre proposed a new implementation of Dedukti [36], based on the Bindlib library for the representation of structures with binders [38]. The libraries generated for Dedukti are compatible with this new implementation, and can be type-checked with minor modifications.

During the first months of his PhD, G. Férey adapted the higher-order pattern matching and convertibility checking algorithms to implemented support for rewriting modulo associative-commutative (AC) symbols in Dedukti.

6.3. Interoperability

F. Thiré has finished to implement a translation of an arithmetic library from Matita to OpenTheory. This work can be decomposed in two steps: A first step goes from Matita to a new logic called STTforall while a second step goes from STTforall to OpenTheory. This translation will be described in two separate papers. The first paper that will be submitted to FSCD 2018 describe the logic STTforall and its translation to HOL while the second paper explains the translation from Matita to STTforall. STTforall is a very simple logic and so, it is easy to translate proofs from this logic to other proofs assistants. For example, a translation from STTforall to Coq has also been implemented by F. Thiré. Two new tools have been implemented to make these translations:

- Dkmeta is a tool that translates terms thanks to the rewrite engine of Dedukti
- Ediloh is a tool that translates terms from STTforall them in OpenTheory

F.Gilbert developed a first prototype for the extraction of proofs from the proof assistant PVS that can be verified externally. The system PVS is based on the dichotomy between a *type-checker* and a *prover*. This proof extraction mechanism is built by instrumenting the PVS *prover*, but does not contain any typing information from the *type-checker* at this stage. Proofs can be built for any PVS theory. However, some reasoning steps rely on unverified assumptions. For a restricted fragment of PVS, the proofs are exported to Dedukti, and the unverified assumptions are proved externally using the automated theorem prover MetiTarski. This work has been published and presented in [15].

6.4. Termination

F. Blanqui revised his paper on "size-based termination of higher-order rewrite systems" submitted to the Journal of Functional Programming [19]. This paper provides a general and modular criterion for the termination of simply-typed λ-calculus extended with function symbols defined by user-defined rewrite rules. Following a work of Hughes, Pareto and Sabry for functions defined with a fixpoint operator and pattern-matching [33], several criteria use typing rules for bounding the height of arguments in function calls. In this paper, we extend this approach to rewriting-based function definitions and more general user-defined notions of size.

R. Lepigre worked on his paper "Practical Subtyping for System F with Sized (Co-)Induction" [39] (joint work with C. Raffalli), which was submitted to the journal Transactions on Programming Languages and Systems (TOPLAS) and is now under revision. This paper proposes a practical type system for a rich, normalizing, extension of (Curry-style) System F. The termination of recursive programs is established using a new mechanism based on circular proofs, which is also used to deal with (sized) inductive and coinductive types (in subtyping). The idea is to build (possibly ill-formed) infinite, circular typing (resp. subtyping) derivations, and to check for their well-foundedness a posteriori. The normalization proof then follows using standard realizability (or reducibility) techniques, the main point being that the adequacy lemma can still be proved by (well-founded) induction on the structure of the "circular" typing (resp. subtyping) derivations.

6.5. Proof theory

G. Burel developed a general framework, focusing with selection, of which various logical systems are instances: ordinary focusing, refinements of resolution, deduction modulo theory, superdeduction and beyond [20]. This strengthens links between sequent calculi and resolution methods.

F. Gilbert developed a constructivization algorithm, taking as input the classical proof of some formula and generating as output, whenever possible, a constructive proof of the same formula. This result has been published and presented in [14].

F. Gilbert submitted his PhD dissertation (work document [25]), centered on the extension of higher-order logic with predicate subtyping. Predicate subtyping is a key feature of the proof assistant PVS, allowing to define types from predicates – for instance, using this feature, the type of even numbers can be defined from the corresponding predicate. The core of this work is the definition of a language of verifiable certificates for predicate subtyping, as well as the proof of two properties of this language: a cut-elimination theorem, a theorem of conservativity over higher-order logic. F. Gilbert presented this language of certificates as well as the cut-elimination theorem at the workshop TYPES 2017.

6.6. Automated theorem proving

G. Bury presented the mSAT library at the OCaml workshop during the International Conference on Functional Programming [21]. This library provides an efficient SAT/SMT solver core written in OCaml, and presented as a functor to allow instantiation with different theories.

6.7. Program verification

R. Lepigre submitted a paper describing the PML_2 programming language and proof assistant [35], which was the main object of his recently defended PhD thesis [37].

6.8. Quantum computing

A. Díaz-Caro and G. Dowek have developed a type system for the λ-calculus that permits to distinguish duplicable terms from non duplicable ones. This work has been presented at Theory and Practice of Natural Computing [13].

GALLIUM Project-Team

7. New Results

7.1. Formal verification of compilers and static analyzers

7.1.1. The CompCert formally-verified compiler

Participants: Xavier Leroy, Daniel Kästner [AbsInt GmbH], Michael Schmidt [AbsInt GmbH], Bernhard Schommer [AbsInt GmbH], Prashanth Mundkur [SRI International].

In the context of our work on compiler verification (see section 3.3.1), since 2005 we have been developing and formally verifying a moderately-optimizing compiler for a large subset of the C programming language, generating assembly code for the ARM, PowerPC, RISC-V and x86 architectures [9]. This compiler comprises a back-end part, translating the Cminor intermediate language to PowerPC assembly and reusable for source languages other than C [8], and a front-end translating the CompCert C subset of C to Cminor. The compiler is mostly written within the specification language of the Coq proof assistant, from which Coq's extraction facility generates executable OCaml code. The compiler comes with a 100000-line, machine-checked Coq proof of semantic preservation establishing that the generated assembly code executes exactly as prescribed by the semantics of the source C program.

This year, we improved the CompCert C compiler in several directions:

- The support for 64-bit target processors that was initiated last year was improved and released as part of version 3.0 of CompCert. CompCert has been supporting 64-bit integer arithmetic since 2013. However, pointers and memory addresses were still assumed to be 32 bits wide. CompCert 3.0 lifts this restriction by parameterizing the compiler over the bit width of memory addresses. This required extensive changes throughout the back-end compiler passes and their correctness proofs.
- The x86 code generator, initially 32-bit only, was extended to handle 64-bit x86 as well. This is the first instantiation of the generic support for 64-bit target architectures mentioned above. This extension greatly improves the usability and performance of CompCert on servers and PCs, where x86 64-bit is the dominant architecture.
- Support for the RISC-V processor architecture was added to CompCert. Prashanth Mundkur contributed a prototype port targeting 32-bit RISC-V. Xavier Leroy extended this port to target 64-bit RISC-V as well and to integrate it in CompCert 3.1. While not commercially available yet, the RISC-V architecture is used in many academic verification projects.
- Several minor optimizations were added to address inefficiencies observed in AbsInt's customer code. The most notable one is the optimization of leaf functions to avoid return address reloads.
- Error and warning messages were improved and made more like those of GCC and Clang. Command-line flags were added to control which warning to emit and which warnings to treat as fatal errors.

We released version 3.0 of CompCert in February 2017 incorporating support for 64-bit architectures, and version 3.1 in August 2017 incorporating the other enhancements listed above.

Two papers describing industrial uses of CompCert for critical software were written, with Daniel Kästner from AbsInt as lead author. The first paper [24] was presented at the 2017 symposium of the British Safety-Critical Systems Club. The second paper [23] will be presented in January 2018 at the ERTS congress. It describes the use of CompCert to compile software for nuclear power plant equipment developed by MTU Friedrichshafen, and the required certification of CompCert according to the IEC 60880 regulations for the nuclear industry.

7.1.2. A verified model of register aliasing in CompCert

Participants: Gergö Barany, Xavier Leroy.

In the setting of the ASSUME ITEA3 project, Gergö Barany and Xavier Leroy are working on implementing a CompCert back-end for the Kalray MPPA processor architecture. This architecture features pervasive register aliasing: each of its 64-bit registers can also be accessed as two separate 32-bit halves. The ARM architecture's floating-point register file is similarly aliased. Modifying a superregister invalidates the data stored in subregisters and vice versa; this behavior was not yet modeled in CompCert's semantics.

Integrating subregister aliasing in CompCert involved re-engineering much of its semantic model of the register file and of the call stack. Rather than simple mappings of locations to values, the register file and the stack are now modeled more realistically as blocks of memory containing bytes that represent fragments of values. In this way, we can verify a semantic model in which a 64-bit register or stack slot may contain either a single 64-bit value or a pair of two unrelated 32-bit values. This ongoing work is nearing completion.

7.1.3. Random program generation for compiler testing

Participant: Gergö Barany.

Randomized testing is a powerful tool for finding bugs in compilers. In a project aimed at finding missed compiler optimizations, Gergö Barany wanted to use such random testing techniques, but found that the standard random C program generator, Csmith, generates very large amounts of dead code. This is code whose results are never used and that can therefore be removed by the compiler.

The presence of large amounts of dead code prevents testing optimizations: almost all of the code is trivially removed by compilers' dead code elimination passes. Gergö resolved this problem by designing a new approach to random program generation. The new generator generates code backwards and performs a simultaneous liveness analysis of the program to rule out the generation of dead code. Its practical evaluation shows that it is much more efficient than Csmith at generating programs that compile to large amounts of machine code with a much more varied instruction mix than Csmith-generated code. In addition, the new generator is much faster than Csmith, because it is designed to work in a single, linear pass, without generating invalid states that cause backtracking. This work resulted in the development of the `ldrgen` tool, and was presented at LOPSTR 2017 [34].

7.1.4. Testing compiler optimizations

Participant: Gergö Barany.

Compilers should be correct, but they should ideally also generate machine code that is as efficient as possible. Gergö Barany started work on adapting compiler correctness testing techniques for testing the quality of the generated code.

In a differential testing approach, one generates random C programs, compiles them with different compilers, then compares the generated code. The comparison is done by a custom binary analysis tool that Gergö developed for this purpose. This tool assigns scores to programs according to various criteria such as the number of instructions, the number of reads from the stack (for comparing the quality of register spilling), or the numbers of various other classes of instructions affected by optimizations of interest. New criteria can be added using a simple plug-in system. If the binaries generated by different compilers are assigned different scores, the input program is considered interesting, and it is reduced to a minimal test case using an off-the-shelf program reducer (C-Reduce).

This automated process often results in small, simple examples of missed optimizations: optimizations that compilers should be able to perform, but that they failed to apply for various reasons. Gergö found previously unreported missing arithmetic optimizations, as well as individual cases of unnecessary register spilling, missed opportunities for register coalescing, dead stores, redundant computations, and missing instruction selection patterns. Several of these missed optimization issues were reported and fixed in the GCC, Clang, and CompCert compilers. An article describing this work is currently under review, and work is in progress on other binary analysis techniques that can find further missed optimizations.

7.1.5. Towards a verified compilation stack for concurrent programs

Participants: Jean-Marie Madiot, Andrew Appel [Princeton University].

The verified compiler CompCert compiles programs from C to assembly while preserving their semantics, thus allowing formal reasoning on source programs, which is much more tractable than reasoning on assembly code. It is however limited to sequential programs, running as one thread on one processor. Jean-Marie Madiot is working to extend CompCert to shared-memory concurrency *and* to provide users with techniques to reason and prove properties about concurrent programs.

Concurrent Separation Logic is used to reason about source programs and prove their correctness with respect to a "concurrent permission machine". The programs are compiled by a concurrency-aware version of CompCert. As of 2017, this has been done for the x86 architecture only.

This project is a continuation of a collaboration with Andrew Appel's team at Princeton University. Appel's team has been working for several years on the "Verified Software Toolchain" project, which provides users with tools to establish properties of sequential programs. Jean-Marie Madiot has been extending the program logic to shared-memory concurrency and developing a new proof of concurrent separation logic that is both formalised and usable in this setting. A paper has been submitted and rejected and is being improved.

Jean-Marie Madiot is now also working on a more general adaptation of CompCert to the reasoning principles of concurrency, and started a collaboration to adapt it to architectures other than x86 (see Section 7.3.4).

7.1.6. Verified compilation of Lustre

Participants: Xavier Leroy, Timothy Bourke [team Parkas], Lélio Brun [team Parkas], Pierre-Évariste Dagand [team Whisper], Marc Pouzet [team Parkas], Lionel Rieg [Yale University].

The Velus project of team Parkas develops a compiler for the Lustre reactive language that generates CompCert Clight intermediate code and is proved correct using the Coq proof assistant. A paper describing the Velus compiler and its verification was presented at the conference PLDI 2017 [20]. Xavier Leroy contributed to the verification of the final pass of Velus, the one that translates from the Obc object-oriented intermediate language of Velus to the Clight C-like, early intermediate language of CompCert. The correctness proof of this pass captures the shape of memory states during execution using formulas from separation logic. The separation logic assertions for CompCert memory states used in this proof come from a library that Xavier Leroy developed last year to help revise the proof of the "stacking" pass of CompCert, and that Timothy Bourke and Xavier Leroy later extended with a "magic wand" operator.

7.2. Language design and type systems

7.2.1. Refactoring with ornaments in ML

Participants: Thomas Williams, Didier Rémy.

Thomas Williams and Didier Rémy continued working on ornaments for program refactoring and program transformation in ML. Ornaments have been introduced as a way of describing changes in data type definitions that preserve the recursive structure but can reorganize, add, or drop pieces of data. After a new data structure has been described as an ornament of an older one, the functions that operate on the bare structure can be partially or sometimes totally lifted into functions that operate on the ornamented structure.

This year, Williams and Rémy continued working on the description of the lifting algorithm: using ornament inference, an ML program is first elaborated into a generic program, which can be seen as a template for all possible liftings of the original program. The generic program is defined in a superset of ML. It can then be instantiated with specific ornaments, and simplified back into an ML program. Williams and Rémy studied the semantics of this intermediate language and used it to prove the correctness of the lifting, using logical relations techniques. A paper has been accepted for presentation at POPL 2018 [14]. A research report gives more technical details [30].

On the practical side, several families of case studies have been explored, including refactoring and code specialization, as so as to make certain existing invariants apparent, or so as to use more efficient data structures. We improved the user interface of the prototype implementation so as to make it easier to write useful examples. We are currently developing a new version of the prototype that will handle most of the OCaml language.

7.3. Shared-memory parallelism

7.3.1. The Linux Kernel Memory Model

Participants: Luc Maranget, Jade Alglave [University College London–Microsoft Research, UK], Paul Mckenney [IBM Corporation], Andrea Parri [Sant'Anna School of Advanced Studies, PISA, Italy], Alan Stern [Harvard University].

Modern multi-core and multi-processor computers do not follow the intuitive "Sequential Consistency" model that would define a concurrent execution as the interleaving of the executions of its constituent threads and that would command instantaneous writes to the shared memory. This situation is due both to in-core optimizations such as speculative and out-of-order execution of instructions, and to the presence of sophisticated (and cooperating) caching devices between processors and memory. Luc Maranget is taking part in an international research effort to define the semantics of the computers of the multi-core era, and more generally of shared-memory parallel devices or languages, with a clear initial focus on devices.

This year saw progress as regards languages. To wit, a two-year effort to define a weak memory model for the Linux Kernel has yielded an article in the *Linux Weekly News* online technical magazine [31], and a scholarly paper accepted for publication at the *Architectural Support for Programming Languages and Operating Systems* (ASPLOS) conference in March 2018. While targeting different audiences, both articles describe a formal model that defines how Linux programs are supposed to behave. The model is of course a CAT model, hence is understood by the **herd** simulator (Section 7.3.3) that allows programmers to experiment and develop an intuition. The model has been tested against hardware and refined in consultation with maintainers. Finally, the ASPLOS article formalizes the *fundamental law of the Read-Copy-Update synchronization mechanism*, and proves that one of its implementations satisfies this law.

For the record, Luc Maranget also co-authored a paper that has been presented at POPL 2017 [22]. This work, which we described last year, is joint work with many researchers, including S. Flur and other members of P. Sewell's team (University of Cambridge) as well as M. Batty (University of Kent). Moreover, Luc Maranget still interacts with the Cambridge team, mostly by providing tests and performing comparisons between his axiomatic models and the operational models developed by this team.

7.3.2. ARMv8 and RISC-V memory models

Participants: Will Deacon [ARM Ltd], Luc Maranget, Jade Alglave [University College London–Microsoft Research, UK].

Jade Alglave and Luc Maranget helped Will Deacon, an engineer at ARM Ltd., who developed a model for the ARMv8 64-bit processor. Will wrote a CAT model, which ARM uses internally as a specification. (CAT is the domain-specific language for describing memory models and is understood by the **herd** simulator; see Section 7.3.3 .) ARM's official documentation presents a natural language transliteration of the CAT model.

Luc Maranget also joined the RISC-V consortium (https://riscv.org/) as an individual and as a member of the memory model group. He takes part in the development of the memory model of this open architecture, mostly by writing CAT models and reviewing tests that will be part of the documentation. A CAT model will be part of the next version (V2.3) of the User-Level ISA Specification.

7.3.3. Improvements to the diy tool suite

Participants: Luc Maranget [**contact**], Jade Alglave [University College London–Microsoft Research, UK].

The **diy** suite (for "Do It Yourself") provides a set of tools for testing shared memory models: the litmus tool for running tests on hardware, various generators for producing tests from concise specifications, and **herd**, a memory model simulator. Tests are small programs written in x86, Power, ARM, generic (LISA) assembler, or a subset of the C language that can thus be generated from concise specifications, run on hardware, or simulated on top of memory models. Test results can be handled and compared using additional tools.

This year's new features are a model for the Linux Kernel developed as a collaborative effort (see Section 7.3.1) and an ongoing RISC-V model transliterated by Luc Maranget from the model elaborated by the RISC-V committee which Luc Maranget joined this year (see Section 7.3.2). Those new models were made possible due to significant extensions of **diy**, such as a new tool chain for RISC-V and the extension of the macro system so as to handle most of the memory-model-related macros used by Linux kernel developers.

7.3.4. Towards formal software verification with respect to weak memory models

Participants: Jean-Marie Madiot, Jade Alglave [University College London & Microsoft Research Cambridge], Simon Castellan [Imperial College London].

Past research efforts on weak memory models have provided both academia and industry with very efficient tools to precisely describe memory models and to carefully test them on a wide variety of architectures. While these models give us a good understanding of complex *hardware* behaviors, exploiting them to formally guarantee the good behavior of *software* remains practically out of reach.

A difficulty is that weak memory models are described in terms of properties of graphs of execution candidates. Because graphs are far from the usual way of defining programming language semantics, because execution candidates are not defined formally, and because existing proofs of "data-race freedom" (DRF) theorems are hard to fathom and formally imprecise, there is a strong demand in the programming language community for a formal account of weak memory models.

In 2017, Jean-Marie Madiot started a collaboration with weak memory model expert Jade Alglave and concurrent game semantics researcher Simon Castellan to tackle these problems. The idea is to have a formal description, using partial-order techniques similar to the ones used in game semantics, of execution candidates. On the other side, a given model of shared memory is then described in terms of partial orders, and the composition of those partial orders provides the final possible executions of a given program in a given architecture. This should yield a formal semantics for programs in a weak memory setting, and should allow proving a DRF theorem so as to connect this semantics to more standard sequentially consistent semantics. A success in this direction would finally allow tractable verification of concurrent programs, particularly in combination with Madiot's ongoing work on a generalization to concurrency of the CompCert certified compiler (see Section 7.1.5).

7.3.5. Granularity control for parallel programs

Participants: Umut Acar, Vitaly Aksenov, Arthur Charguéraud, Adrien Guatto, Mike Rainey, Filip Sieczkowski.

The DeepSea team focused this year on the development of techniques for controlling granularity in parallel programs. Granularity control is an essential problem because creating too many tasks may induce overwhelming overheads, while creating too few tasks may harm the ability to process tasks in parallel. Granularity control turns out to be especially challenging for nested parallel programs, i.e., programs in which parallel constructs such as fork-join or parallel-loops can be arbitrarily nested. Two different approaches were investigated.

The first approach is based on the use of asymptotic complexity functions provided by the programmer, combined with runtime measurements to estimate the constant factors that apply. Combining these two sources of information allows to predict with reasonable accuracy the execution time of tasks. Such predictions may be used to guide the generation of tasks, by sequentializing computations of sufficiently-small size. An analysis is developed, establishing that task creation overheads are indeed bounded to a small fraction of the total runtime. These results builds upon prior work by the same authors [39], extending it with a carefully-designed algorithm for ensuring convergence of the estimation of the constant factors deduced from the measures, even in the face of noise and cache effects, which are taken into account in the analysis. The approach is demonstrated on a range of benchmarks taken from the state-of-the-art PBBS benchmark suite. A paper describing the results is under preparation.

The second approach is based on an instrumentation of the runtime system. The idea is to process parallel function calls just like normal function calls, by pushing a frame on the stack, and only subsequently promoting these frames as threads that might get scheduled on other cores. The promotion of frames takes place at regular time intervals, which is why we named this approach *heartbeat scheduling*. Unlike prior approaches such as *lazy scheduling*, in which promotion is guided by the workload of the system, heartbeat scheduling can be proved to induce only small scheduling overheads, and to not asymptotically reduce the amount of parallelism inherent in the program. The theory behind the approach is formalized in Coq. It is also implemented through instrumented C++ programs, and evaluated on PBBS benchmarks. A paper describing this approach was submitted to an international conference.

7.3.6. Non-zero indicators: a provably-efficient, concurrent data structure

Participants: Umut Acar, Mike Rainey.

This work, conducted in collaboration with Naama Ben David from Carnegie Mellon University, investigates the design and analysis of an implementation of a concurrent data structure called *non-zero indicator*. This data structure plays a crucial role in the scheduling of nested parallel programs: it is used to handle dependency resolution among parallel tasks. Concretely, a non-zero indicator is initialized with value 1, and it supports the following two concurrent operations, which may be invoked by threads that have knowledge that the counter is non-zero: (1) atomically increase the counter by one unit, and (2) atomically decrease the counter by one unit, and detect whether the counter reaches zero. While a trivial implementation can be set up using an atomic operation on a shared memory cell (e.g., fetch-and-add), the key challenge is to design a non-zero indicator that scales well to hundreds if not thousands of threads, without suffering from contention.

Prior work leverages dynamic tree data structures to tame contention [42]. Yet, such prior work, as well as most concurrent data structures in general, are analyzed empirically, omitting asymptotic bounds on their efficiency. In this work, we propose a new variant of a tree-based non-zero indicator implementation, for which we are able to present a formal analysis establishing bounds on the worst-case contention of concurrent updates. Our analysis is the first to achieve relevant bounds of this kind. Furthermore, we demonstrate in practice that our proposal improves scalability, compared with a naive fetch-and-add atomic counter, and also compared with the original tree-based data structure. Our work was presented at PPoPP [16].

7.3.7. Efficient sequence data structures for ML

Participants: Arthur Charguéraud, Mike Rainey.

The use of sequence containers, including stacks, queues, and double-ended queues, is ubiquitous in programming. When the maximal number of elements to be stored is not known in advance, containers need to grow dynamically. For this purpose, most ML programs rely on either lists or vectors. These data structures are inefficient, both in terms of time and space usage. In this work, we investigate the use of data structures based on *chunks*, adapting ideas from some of our prior work implemented in C++ [38]. Each chunk stores items in a fixed-capacity array. All chunks are linked together to represent the full sequence. These chunk-based structures save a lot of memory and generally deliver better performance than classic container data structures for long sequences. We measured a 2x speedup compared with vectors, and up to a 3x speedup compared with long lists. This work was presented at the ML Family Workshop [36]. Generalization of this work to double-ended sequences and to persistent sequences is under progress.

7.3.8. A parallel algorithm for the dynamic trees problem

Participants: Umut Acar, Vitaly Aksenov.

Dynamic algorithms are used to compute a property of some data while the data undergoes changes over time. Many dynamic algorithms have been proposed, but nearly all of them are sequential.

In collaboration with Sam Westrick (Carnegie Mellon University), Umut Acar and Vitaly Aksenov investigated the design of an efficient parallel dynamic tree data structure. This data structure supports four operations, namely insertion and deletion of vertices and edges; these operations can be executed in parallel. The proposed data structure is work-efficient and highly parallel. A preliminary version of this work was presented in a brief announcement at SPAA 2017 [15].

7.3.9. A concurrency-optimal binary search tree

Participant: Vitaly Aksenov.

In joint work with Vincent Gramoli (IT School of Information Technologies, Sydney), Petr Kuznetsov (Telecom ParisTech), Anna Malova (Washington University in St Louis), and Srivatsan Ravi (Purdue University), Vitaly Aksenov proposed a concurrency-optimal implementation of binary search trees. Concurrency-optimality means that the data structure allows all interleavings of the underlying sequential implementation, except those that would violate linearizability. Aksenov and co-authors show that none of the state-of-the-art concurrent binary search trees are concurrency-optimal, and they experimentally verify that the new concurrency-optimal binary search tree is competitive with known implementations. This work was presented at Euro-Par 2017 [17].

7.4. The OCaml language and system

7.4.1. The OCaml system

Participants: Damien Doligez, Xavier Leroy, Luc Maranget, David Allsop [Cambridge University], Florian Angeletti, Alain Frisch [Lexifi], Jacques Garrigue [University of Nagoya], Sébastien Hinderer [SED], Nicolás Ojeda Bär [Lexifi], Thomas Refis [Jane Street], Gabriel Scherer [team Parsifal], Mark Shinwell [Jane Street], Leo White [Jane Street], Jeremy Yallop [Cambridge University].

This year, we released four versions of the OCaml system: versions 4.04.1 and 4.04.2 are minor releases that fix about 16 issues; versions 4.05.0 and 4.06.0 are major releases that introduce some new features, many improvements in usability and performance, and fix about 100 issues. The most important new features are:

- Character strings are now immutable (read-only) by default. This completes the evolution of OCaml towards immutable strings that started in 2014 with the introduction of a compile-time option to separate text-like read-only strings from array-like read-write byte sequences. This option is now the default, making OCaml programs safer and clearer.
- Extensions of the "destructive substitution" operator over module signatures (*sig* `with type t :=` ...) to make it more general and more widely usable.
- Support for the UTF8 encoding of Unicode characters in strings was improved with the introduction of an escape `\u{XXXX}` in string literals, and more importantly with a complete overhaul of the OCaml interface for Windows system calls that make them compatible with UTF8-encoded Unicode.
- An alternate register allocator based on linear scan was added and can be selected to reduce compilation times.

On the organization side, we switched to a deadline-based release cycle whereby a major release occurs at a set date with the features that are ready by that date, instead of waiting for a set of new features to be ready. Releases 4.05.0 and 4.06.0 were produced in this manner at 6-months intervals. Damien Doligez and Gabriel Scherer served as release managers.

Sébastien Hinderer worked on integrating `ocamltest`, the compiler's test driver he developed last year, in the 4.06 release of OCaml. He migrated a large part of the test suite from the former Makefile-based infrastructure to `ocamltest`. He also started to rewrite OCaml's build system so that the compiler can be built in parallel as much as its dependencies allow.

We have improved our Continuous Integration infrastructure by taking advantage of Jenkins features such as configuration matrices, adding five new architectures (ARM-64, Fedora, FreeBSD, PPC64-LE, Ubuntu), and upgrading to the latest version of MacOS. Our testing is now done on all of the major architectures that are officially supported by OCaml.

7.4.2. Type-checking the OCaml intermediate languages

Participants: Pierrick Couderc [ENSTA-ParisTech & OCamlPro], Grégoire Henry [OCamlPro], Fabrice Le Fessant, Michel Mauny.

This work aims at designing and implementing a consistency checker for the type-annotated abstract syntax trees (TASTs) produced by the OCaml compiler. When presented as inference rules, the different cases of this TAST checker can be read as the rules of the OCaml type system. Proving the correctness of (part of) the checker would prove the soundness of the corresponding part of the OCaml type system. A preliminary report on this work has been presented at the 17th Symposium on Trends in Functional Programming (TFP 2016).

In 2017, Pierrick Couderc formalized the consistency checker, and wrote a Coq proof of its correctness. The dissertation is being written, and Pierrick's Ph.D. defense should take place at the beginning of 2018.

7.4.3. Optimizing OCaml for satisfiability problems

Participants: Sylvain Conchon [LRI, Univ. Paris Sud], Albin Coquereau [ENSTA-ParisTech], Mohamed Iguernelala [OCamlPro], Fabrice Le Fessant, Michel Mauny.

This work aims at improving the performance of the Alt-Ergo SMT solver, implemented in OCaml. For safety reasons and to ease reasoning about its algorithms, the implementation of Alt-Ergo uses as much as possible a functional programming style and persistent data structures, which are sometimes less efficient than imperative style and mutable data. Moreover, some efficient algorithms, such as CDCL SAT solvers, are naturally expressed in an imperative style.

We therefore explored the replacement of Alt-Ergo's default, functional, SAT solver by an imperative CDCL solver. In a first step, we reimplemented a C++ version of miniSAT in OCaml. A comparison of their respective performance showed that the OCaml version is slower and has more cache misses.

In a second step, we studied the use of the imperative miniSAT-like SAT solver in Alt-Ergo. The integration is actually not immediate because of the interaction between this solver and both the theories and the quantifier instantiation engines of Alt-Ergo. In fact, although the default (functional) SAT solver of Alt-Ergo is not as effective as a CDCL solver for reasoning on pure Boolean problems, its smart interaction with theories and instantation engines makes it quite effective in the context of program verification.

7.4.4. Type compatibility checking for dynamically-loaded OCaml data

Participants: Florent Balestrieri [ENSTA-ParisTech], Michel Mauny.

The SecureOCaml project (FUI 18) aims at enhancing the OCaml language and environment in order to make it more suitable for building secure applications, following the recommendations published by the French ANSSI in 2013. Florent Balestrieri (ENSTA-ParisTech) represents ENSTA-Paristech in this project for 2016 and 2017.

The first year has been dedicated to designing and producing an effective OCaml implementation that checks whether a memory graph – typically the result obtained by unmarshalling some data – is compatible with a given OCaml type, following the algorithm designed by Henry *et al.* in 2012. Because the algorithm requires a runtime representation of OCaml types, Florent Balestrieri implemented a library for generic programming in OCaml. This library was presented at the OCaml Users and Developers Workshop in 2016 [40]; an extended version of this paper has been submitted [33]. He also implemented a type-checker which, when given a type and a memory graph, checks whether the former could be the type of the latter. In 2017, Florent Balestrieri implemented a prototype type-checker for OCaml bytecode.

7.4.5. Visitors

Participant: François Pottier.

Traversing and transforming abstract syntax trees that involve name binding is notoriously difficult to do in a correct, concise, modular, customizable manner. In 2017, François Pottier addressed this problem in the setting of OCaml by proposing visitor classes as partial, composable descriptions of the operations that one wishes to perform on abstract syntax trees. By combining auto-generated visitor classes (which have no knowledge of binding) and hand-written visitor classes (each of which knows about a specific binding construct, a specific representation of names, and/or a specific operation on abstract syntax trees), a wide range of operations can be defined. A syntax extension for OCaml has been released under the name `visitors` and this work has been presented at the conference ICFP 2017 [13].

7.4.6. Improvements in Menhir

Participant: François Pottier.

In 2017, François Pottier incorporated several improvements, proposed by Frédéric Bour, to the Menhir parser generator. Many functions were added to Menhir's incremental API, which (at runtime) allows inspecting and updating the parser's state from the outside. A new library, MENHIRSDK, was introduced, which (at compile-time) allows inspecting the grammar and the automaton constructed by Menhir. Together, these improvements allow new features to be programmed outside of Menhir; the advanced error recovery mode implemented in the Merlin IDE is an example.

François Pottier also improved the termination test that takes place before parameterized symbols are expanded away. The new test, it is hoped, should reject the grammar if and only if expansion would not terminate. This improves the expressive power of the grammar description language.

7.5. Software specification and verification

7.5.1. Formal reasoning about asymptotic complexity

Participants: Armaël Guéneau, Arthur Charguéraud, François Pottier.

For several years, Arthur Charguéraud and François Pottier have been investigating the use of Separation Logic, extended with Time Credits, as an approach to the formal verification of the time complexity of OCaml programs. An extended version of their work on the UnionFind algorithm has appeared in the *Journal of Automated Reasoning* [11]. In this work, the complexity bounds that are established involve explicit constants: for instance, the complexity of *find* is $2\alpha(n) + 4$.

Armaël Guéneau, who is supervised by Arthur Charguéraud and François Pottier, is working on relaxing this approach so as to use asymptotic bounds: e.g., the advertised complexity of *find* should be $O(\alpha(n))$. The challenge is to give a formal account of the O notation and of its properties and to develop techniques that make asymptotic reasoning as convenient in Coq as it seemingly is on paper.

For that purpose, this year, Armaël Guéneau developed two Coq libraries. A first library gives a formal definition of the O notation, provides proofs for many commonly used lemmas, as well as a number of tactics that automate the application of these lemmas. A second library implements a simple yet very useful mechanism, allowing the user to delay and collect proof obligations in Coq scripts. Using these libraries, Armaël extended the CFML tool with support for making asymptotic time complexity claims as part of specifications. He developed tactics that perform (guided) inference and resolution of recursive equations for the cost of recursive programs.

Armaël evaluated this framework on several small-scale case studies, namely simple algorithms such as binary search, selection sort, and the Bellman-Ford algorithm. This work has been accepted for publication at the conference ESOP 2018.

7.5.2. Revisiting the CPS transformation and its implementation

Participant: François Pottier.

While preparing an MPRI lecture on the CPS transformation, François Pottier did a machine-checked proof of semantic correctness for Danvy and Filinski's properly tail-recursive, one-pass, call-by-value CPS transformation.

He proposed a new first-order, one-pass, compositional formulation of the transformation. He pointed out that Danvy and Filinski's simulation diagram does not hold in the presence of `let` and proved a slightly more complex diagram, which involves parallel reduction. He suggested representing variables as de Bruijn indices and showed that, thanks to state-of-the-art libraries such as Autosubst, this does not represent a significant impediment to formalization. Finally, he noted that, given this representation of terms, it is not obvious how to efficiently implement the transformation. To address this issue, he proposed a novel higher-order formulation of the CPS transformation, proved that it is correct, and informally argued that it runs in time $O(n \log n)$.

This work has been submitted for publication in a journal.

7.5.3. Zenon

Participant: Damien Doligez.

This year, Damien Doligez did maintenance work on Zenon: updating to the latest version of OCaml and fixing a few bugs. He also started work on adding a few minor features, such as inductive proofs for mutually inductive types.

7.5.4. TLA+

Participants: Damien Doligez, Leslie Lamport [Microsoft Research], Martin Riener [team VeriDis], Stephan Merz [team VeriDis].

Damien Doligez is head of the "Tools for Proofs" team in the Microsoft-Inria Joint Centre. The aim of this project is to extend the TLA+ language with a formal language for hierarchical proofs, formalizing Lamport's ideas [44], and to build tools for writing TLA+ specifications and mechanically checking the proofs.

Damien is still working on a new version of TLAPS and has started writing a formal description of the semantics of TLA+.

MARELLE Project-Team

6. New Results

6.1. Implementing Theorem Proving in Higher Order Logic Programming

Participants: Enrico Tassi, Luc Chabassier, Cyril Cohen, Cvetan Dunchev [University of Bologna], Ferruccio Guidi [University of Bologna], Claudio Sacerdoti Coen [University of Bologna].

We are designing a Coq plugin named elpi providing an extension language based on λ-prolog to write new commands and tactics. This year, we re-designed the constraint handling engine of the elpi interpreter. Luc Chabassier illustrated the use of this extension on the problem of generating automatically equality test functions for arbitrary recursive types, together with their proof of correctness.

Another experiment was conducted by Cyril Cohen on using elpi to compute genericity theorems. For now the unary and binary cases have been covered in a concise fashion.

An article on this topic has been submitted to MSCS [19], a presentation will also be given at the CoqPL workshop [21].

6.2. Coqoon: An IDE for interactive proof development in Coq

Participants: Enrico Tassi, Alexander Faithfull [ITU Copenhagen], Jesper Bengtson [ITU Copenhagen], Carst Tankink.

The work of previous years on Coqoon has been published in an international journal [6].

6.3. Proofs of transcendence

Participants: Sophie Bernard, Yves Bertot, Laurence Rideau.

Sophie Bernard completed a proof of the Lindemann-Weierstrass theorem concerning the algebraic independence of spans of exponentials of rationally dependent numbers. This result required that we extend the theory of symmetric multivariate polynomials in order to formalize the notion of conjugates of a polynomial. This was described in an article presented at an international conference [13] and at a workshop associated to ANR project FastRelax.

6.4. Cubical type theory and univalent foundations

Participants: Cyril Cohen, Anders Mörtberg, Benedikt Ahrens [ASCOLA project-team, Inria and LINA Nantes], Mark Bickford [Cornell Unversity, USA], Thierry Coquand [Chalmers and Göteborg University, Sweden], Simon Huber [Chalmers University, Sweden], Ralph Matthes [CNRS, University of Toulouse].

This work mainly concerns Univalent Foundations and Homotopy Type Theory, especially in the form of cubical type theory. The code is visible at https://github.com/mortberg/cubicaltt. This year, Anders Mörtberg has been working on formalizing cubical set models in univalent type theory and on extending cubical type theory with a general class of higher inductive types, in collaboration with Cyril Cohen, Thierry Coquand and Simon Huber.

Anders Mörtberg extended work with Ralph Matthes, Benedikt Ahrens and Vladimir Voevodsky on the representation of syntax of programming languages using category theory in univalent type theory. This paper was accepted for publication in JAR.

Anders Mörtberg also prepared a series of lectures introducing to cubical type theory. this lead to invited talks at the workshops "Type Theory based Tools (TTT)", and "Syntax and Semantics of Type Theory".

6.5. Formal study of double-word arithmetic algorithms

Participants: Laurence Rideau, Erik Martin-Dorel [IRIT Toulouse], Jean-Michel Muller [CNRS and ENS Lyon], Valentina Popescu [CNRS and ENS Lyon].

As part of the ANR Fastrelax project, we have started to formalize double-word arithmetic algorithms, in particular the sum of a double-word and a floating point number and the sum of two double-word numbers described in the article " Tight and rigourous error bounds for basic building blocks of double-word arithmetic" [24]. The formalization is progressing. A notable event is that we detected a small error in the article proof, which required a correction by the authors.

6.6. Formal study of comparisons between numbers in different formats

Participants: Laurent Théry, Arthur Blot, Jean-Michel Muller [CNRS and ENS Lyon].

We show how a library of formalized mathematics about continuous functions can be used to derive an algorithm that compares two floating point number one in base 2 and one in base 10 [14].

6.7. A formal study of the towers of Hanoi

Participant: Laurent Théry.

The towers of Hanoi is a classical example that illustrates the power of recursive programming. Proving that the recursive program solves the problem is elementary but proving that it is a minimal solver is harder. This is even more difficult if we consider the general problem that considers arbitrary starting and final positions. We present the formalisation of this problem in the Mathematical Component Library [22].

6.8. Formaly study of algorithms to compute π

Participants: Yves Bertot, Laurence Rideau, Laurent Théry.

We studied formal proofs for several algorithms used to computed π to very high precisions, the famous BBP formula and an algorithm derived from it and another algorithm based on arithmetic-geometric means that is used in the MPFR library. These results show that Coq can be used directly to compute a million decimals or the billionth hexa-decimal in isolation [5].

6.9. Formal foundations of 3D geometry for robot manipulators

Participants: Cyril Cohen, Reynald Affeldt [AIST, Japan].

We resumed our collaboration with the team at AIST for the formal description of robotics aspects [7]. Reynald Affeldt visited Sophia Antipolis for 10 days during which we improved the connection between our library for algebra and the Coquelicot library for analysis.

6.10. Formalization of Analysis concepts

Participants: Cyril Cohen, Damien Rouhling.

To study problems in control, we worked on the notion of compacts and showed how to express it using filters, as in Coquelicot.

We experimented with sets of notations to make computing with limits simpler. We also generalized the notion of "big enough" that can usually be found when reasoning about functions at infinity (or sequences) so that it now works with arbitrary filters. Finally, we started experimenting with a new point of view on "small o" notations.

We also started work on formalizing in Coq the Cauchy-Lipschitz theorem (also known as Picard-Lindelöf), which proves the existence and uniqueness of solutions to differential equations.

We expect all these small advances to prepare the ground for work on various aspects of robotics and control. Part of this work was published in an international conference [20].

6.11. Formalization of proofs in control theory

Participants: Damien Rouhling, Cyril Cohen.

We worked on dynamical systems and differential equations. Damien Rouhling fully formalized in Coq LaSalle's invariance principle with the help of Cyril Cohen. This principle uses Lyapunov functions to prove the stability of a dynamical system defined by a differential equation. We wrote a paper about this formalization, which has been published in the proceedings of the ITP 2017 conference [15].

We improved this formalization to apply this principle to an example of robotics and control theory. We formalized in Coq the correctness of a control function for an inverted pendulum. Damien Rouhling wrote a paper about this, accepted for publication at an international conference in early 2018 [20].

6.12. Formalization of graph algorithms

Participants: Yves Bertot, Cyril Cohen, Ran Chen [Chinese Academy of Science], Jean-Jacques Lévy [Pi.r2 and Chinese Academy of Science], Clément Sartori, Laurent Théry.

We studied algorithms to compute strongly connected components in graphs, as a way to prepare a comparative study with the work of Levy and Chen: "A Semi-Automatic Proof of Strong Connectivity" [23].

In a similar vein, Yves Bertot and Clément Sartori have been studying the combinatorial aspects of triangulations, and in particular Delaunay triangulations, seen as graphs. In the long run, we expect this effort to contribute to formal descriptions of Voronoi diagrams and uses in robot motion planning.

6.13. Extension of the CoqEAL library

Participants: Cyril Cohen, Enrico Tassi.

The CoqEAL library provides a framework to connect efficient executable functional programs to the algorithms that are described formally using the mathematical components library. Key aspects rely on the capacity to refine abstract views of the algorithms and data into concrete views, where the efficiency can be fine-tuned. For this refinement, we also need to rely on properties of programming languages such as parametricity. We experimented on relying on the ELPI plugin to implement this parametricity feature. In the long run, this means that the ELPI plugin should play an instrumental role in making CoqEAL easy to use and to extend.

6.14. Formalizing Exterior Algebras

Participants: Maxime Bombar, Cyril Cohen.

We formalized exterior algebras as vector spaces with dimension 2^n. This provides an alternative representation to that constructed earlier by Laurent Théry and Laurent Fuchs. The new representation is closer to the objects found in the mathematical components library.

6.15. Formalizing Cylindrical Algebraic Decomposition

Participants: Boris Djalal, Yves Bertot, Cyril Cohen.

Our study of cylindrical algebraic decomposition requires that we find a good representation of semi-algebraic sets, which are usually determined by a collections of comparisons between polynomial formulas. We wrote an article on this topic, which has been accepted for publication at an international conference to be held in early 2018 [18].

6.16. Formal study of probabilistic programs

Participants: Benjamin Grégoire, Gilles Barthe [IMDEA], François Dupressoir [University of Surrey], Thomas Espitau [UPMC Paris 6], Sebastian Faust [Ruhr Universitat Bochum], Justin Hsu [University of Pennsylvania], Vitor Pereira [INESC TEC], François-Xavier Standaert [Université Catholique de Louvain].

This year, we proposed new logics to make a link between "probabilistic Relationnal Hoare Logic" and the traditionnal notion of couplings from probability theory [12]. We have also showed that coupling can be use to prove non-relationnal properties like uniformity and probabilistic independence [11].

We used Easycrypt to prove the security of Secure Function Evaluation (SFE) based on garble circuits [9].

6.17. Generating Efficient Resistant Code

Participants: Benjamin Grégoire, José Bacelar Almeida [INESC TEC], Manuel Barbosa [INESC TEC], Gilles Barthe [IMDEA], Arthur Blot [ENS Lyon], Vincent Laporte [IMDEA], Tiago Oliveira [INESC TEC], Hugo Pacheco [INESC TEC], Benedikt Schmidt [Google Inc.], Pierre-Yves Strub [Ecole Polytechnique].

We develop a certified compiler named Jasmin to generate high-speed and high-assurance cryptographic code.

Differential power analysis (DPA) is a side-channel attack in which an adversary retrieves cryptographic material by measuring and analyzing the power consumption of the device on which the cryptographic algorithm under attack executes. We introduced new notions/models allowing to check the correctness of counter measures (masking schemes) [10].

6.18. Formal Security Proof in EasyCrypt: case studies and extensions

Participants: Cécile Baritel-Ruet, Benjamin Grégoire.

We completed a formal proof of security for CMAC, a scheme for cipher-based message authentication code. A publication is being submitted on this topic. We also experimented on a formal study of the forking lemma, which is present in many security proofs for signing schemes that rely on lattice problems.

The lessons derived from these experiments lead us to proposing new tools for matching instructions and unifying formulas with meta-variables in EasyCrypt.

6.19. Formalizing Bourbaki-style mathematics

Participant: José Grimm.

Most of the work described here is inspired by the experiment of giving formal proofs in Coq of the exercises found in Bourbaki's exposition of set theory. However, some of the results go beyond what can be found in Bourbaki.

We studied order relations by proving several properties about the *length* and *width* of order relations, for instance showing that when a set has $nm + 1$ elements, the lengh or the with of any order on this set is larger than either n or m. We then considered similar theorems on the set of all parts of a given set, ordered by inclusion. In particular, this gives formal proofs of results by Dilworth and Erdös and Zserkeres.

We also studied ordinal addition, which is non-commutative. Given a finite sequence of ordinals, one can compute the number of different results of the sum of these elements, depending on the order in which this sequence is taken. There is an explicit formula for this number, with a proof that we formalized.

Last, we studied a footnote from Bourbaki, that indicates that 1 is a notation for a term whose normal form has several tens of thousands of signs. We compute this size (about 10^{13} or 10^{60} depending on whether some constructs are given by axioms or by definitions) and provide statistics on the distributions of signs in the normal form.

MEXICO Project-Team

7. New Results

7.1. Optimal constructions for active diagnosis

Published in [4].

Diagnosis is the task of detecting fault occurrences in a partially observed system. Depending on the possible observations, a discrete-event system may be diagnosable or not. Active diagnosis aims at controlling the system to render it diagnosable. Past research has proposed solutions for this problem, but their complexity remains to be improved. Here, we solve the decision and synthesis problems for active diagnosability, proving that (1) our procedures are optimal with respect to computational complexity, and (2) the memory required for our diagnoser is minimal. We then study the delay between a fault occurrence and its detection by the diagnoser. We construct a memory-optimal diagnoser whose delay is at most twice the minimal delay, whereas the memory required to achieve optimal delay may be highly greater. We also provide a solution for parametrized active diagnosis, where we automatically construct the most permissive controller respecting a given delay.

7.2. Diagnosability of Repairable Faults

Published in [3].

The diagnosis problem for discrete event systems consists in deciding whether some fault event occurred or not in the system, given partial observations on the run of that system. Diagnosability checks whether a correct diagnosis can be issued in bounded time after a fault, for all faulty runs of that system. This problem appeared two decades ago and numerous facets of it have been explored, mostly for permanent faults. It is known for example that diagnosability of a system can be checked in polynomial time, while the construction of a diagnoser is exponential. The present paper examines the case of transient faults, that can appear and be repaired. Diagnosability in this setting means that the occurrence of a fault should always be detected in bounded time, but also before the fault is repaired, in order to prepare for the detection of the next fault or to take corrective measures while they are needed. Checking this notion of diagnosability is proved to be PSPACE-complete. It is also shown that faults can be reliably counted provided the system is diagnosable for faults and for repairs.

7.3. Diagnostic et contrôle de la dégradation des systèmes probabilistes

Published in [18].

Le diagnostic actif est opéré par un contrôleur en vue de rendre un système diagnosticable. Afin d'éviter que le contrôleur ne dégrade trop fortement le système, on lui affecte généralement un second objectif en termes de qualité de service. Dans le cadre des systèmes probabilistes, une spécification possible consiste à assurer une probabilité positive qu'une exécution infinie soit correcte, ce qu'on appelle le diagnostic actif sûr. Nous introduisons ici deux spécifications alternatives. La gamma-correction du système affecte à une exécution une valeur de correction dépendant d'un facteur de décote gamma et le contrôleur doit assurer une valeur moyenne supérieure à un seuil fixé. La alpha-dégradation requiert qu'asymptotiquement, à chaque unité de temps une proportion supérieure à alpha des exécutions jusqu'alors correctes le demeure. D'un point de vue sémantique, nous explicitons des liens significatifs entre les différentes notions. Algorithmiquement, nous établissons la frontière entre décidabilité et indécidabilité des problèmes et dans le cas positif nous exhibons la complexité précise ainsi qu'une synthèse, potentiellement à mémoire infinie.

7.4. The Complexity of Diagnosability and Opacity Verification for Petri Nets

Published in [7].

Diagnosability and opacity are two well-studied problems in discrete-event systems. We revisit these two problems with respect to expressiveness and complexity issues. We first relate different notions of diagnosability and opacity. We consider in particular fairness issues and extend the definition of Germanos et al. [ACM TECS, 2015] of weakly fair diagnosability for safe Petri nets to general Petri nets and to opacity questions. Second, we provide a global picture of complexity results for the verification of diagnosability and opacity. We show that diagnosability is NL-complete for finite state systems, PSPACE-complete for safe Petri nets (even with fairness), and EXPSPACE-complete for general Petri nets without fairness, while non diagnosability is inter-reducible with reachability when fault events are not weakly fair. Opacity is ESPACE-complete for safe Petri nets (even with fairness) and undecidable for general Petri nets already without fairness.

7.5. Probabilistic Disclosure: Maximisation vs. Minimisation

Published in [8].

We consider opacity questions where an observation function provides to an external attacker a view of the states along executions and secret executions are those visiting some state from a fixed subset. Disclosure occurs when the observer can deduce from a finite observation that the execution is secret, the ε-disclosure variant corresponding to the execution being secret with probability greater than 1 - ε. In a probabilistic and non deterministic setting, where an internal agent can choose between actions, there are two points of view, depending on the status of this agent: the successive choices can either help the attacker trying to disclose the secret, if the system has been corrupted, or they can prevent disclosure as much as possible if these choices are part of the system design. In the former situation, corresponding to a worst case, the disclosure value is the supremum over the strategies of the probability to disclose the secret (maximisation), whereas in the latter case, the disclosure is the infimum (minimisation). We address quantitative problems (comparing the optimal value with a threshold) and qualitative ones (when the threshold is zero or one) related to both forms of disclosure for a fixed or finite horizon. For all problems, we characterise their decidability status and their complexity. We discover a surprising asymmetry: on the one hand optimal strategies may be chosen among deterministic ones in maximisation problems, while it is not the case for minimisation. On the other hand, for the questions addressed here, more minimisation problems than maximisation ones are decidable.

7.6. D-SPACES: Implementing Declarative Semantics for Spatially Structured Information

Published in [13].

We introduce in this paper D-SPACES, an implementation of constraint systems with space and extrusion operators. Constraint systems are algebraic models that allow for a semantic language-like representation of information in systems where the concept of space is a primary structural feature. We give this information mainly an epistemic interpretation and consider various agents as entities acting upon it. D-SPACES is coded as a c++11 library providing implementations for constraint systems, space functions and extrusion functions. The interfaces to access each implementation are minimal and thoroughly documented. D-SPACES also provides property-checking methods as well as an implementation of a specific type of constraint systems (a boolean algebra). This last implementation serves as an entry point for quick access and proof of concept when using these models. Furthermore, we offer an illustrative example in the form of a small social network where users post their beliefs and utter their opinions.

7.7. Unbounded product-form Petri nets

Published in [14].

Computing steady-state distributions in infinite-state stochastic systems is in general a very difficult task. Product-form Petri nets are those Petri nets for which the steady-state distribution can be described as a natural product corresponding, up to a normalising constant, to an exponentiation of the markings. However, even though some classes of nets are known to have a product-form distribution, computing the normalising constant can be hard. The class of (closed) Π^3-nets has been proposed in an earlier work, for which it is shown that one can compute the steady-state distribution efficiently. However these nets are bounded. In this paper, we generalise queuing Markovian networks and closed Π^3-nets to obtain the class of open Π^3-nets, that generate infinite-state systems. We show interesting properties of these nets: (1) we prove that liveness can be decided in polynomial time, and that reachability in live Π^3-nets can be decided in polynomial time; (2) we show that we can decide ergodicity of such nets in polynomial time as well; (3) we provide a pseudo-polynomial time algorithm to compute the normalising constant.

7.8. Statistical Model-Checking for Autonomous Vehicle Safety Validation

Published in [20].

We present an application of statistical model-checking to the verification of an autonomous vehicle controller. Our goal is to check safety properties in various traffic situations. More specifically, we focus on a traffic jam situation.

The controller is specified by a C++ program. Using sensors, it registers positions and velocities of nearby vehicles and modifies the position and velocity of the controlled vehicle to avoid collisions. We model the environment using a stochastic high level Petri net, where random behaviors of other vehicles can be described. We use HASL, a quantitative variant of linear temporal logic, to express the desired properties. A large family of performance indicators can be specified in HASL and we target in particular the expectation of travelled distance or the collision probability.

We evaluate the properties of this model using COSMOS1. This simulation tool implements numerous statistical techniques such as sequential hypothesis testing and most confidence range computation methods. Its efficiency allowed us to conduct several experiments with success.

7.9. Une sémantique formelle pour les modèles Simulink

Published in [19].

De nombreux projets industriels, notamment dans la construction automobile, font appel à la suite d'outils Simulink pour la conception et la validation de composants critiques représentant des systèmes hybrides c'est-à-dire combinant des aspects discrets et continus. Cependant les formalismes associés ne disposent pas d'une sémantique formelle ce qui peut diminuer la confiance des ingénieurs vis-à-vis des résultats produits. Nous proposons ici une telle sémantique en procédant en deux étapes. Nous développons d'abord une sémantique exacte mais non exécutable. Puis nous l'enrichissons d'une sémantique opérationnelle approchée avec pour objectif une quantification de l'erreur issue de cette approximation.

7.10. The Logical View on Continuous Petri Nets

Published in [5].

Continuous Petri nets are a relaxation of classical discrete Petri nets in which transitions can be fired a fractional number of times, and consequently places may contain a fractional number of tokens. Such continuous Petri nets are an appealing object to study since they over approximate the set of reachable configurations of their discrete counterparts, and their reachability problem is known to be decidable in polynomial time. The starting point of this paper is to show that the reachability relation for continuous Petri nets is definable by a sentence of linear size in the existential theory of the rationals with addition and order. Using this characterization, we obtain decidability and complexity results for a number of classical decision problems for continuous Petri nets. In particular, we settle the open problem about the precise complexity of reachability set inclusion. Finally, we show how continuous Petri nets can be incorporated inside the classical

backward coverability algorithm for discrete Petri nets as a pruning heuristic in order to tackle the symbolic state explosion problem. The cornerstone of the approach we present is that our logical characterization enables us to leverage the power of modern SMT-solvers in order to yield a highly performant and robust decision procedure for coverability in Petri nets. We demonstrate the applicability of our approach on a set of standard benchmarks from the literature.

7.11. Memoryless Determinacy of Finite Parity Games: Another Simple Proof

Published in [24].

Memoryless determinacy of (infinite) parity games is an important result with numerous applications. It was first independently established by Emerson and Jutla [1] and Mostowski [2] but their proofs involve elaborate developments. The elegant and simpler proof of Zielonka [3] still requires a nested induction on the finite number of priorities and on ordinals for sets of vertices. There are other proofs for finite games like the one of Bjørklund, Sandberg and Vorobyovin [4] that relies on relating infinite and finite duration games. We present here another simple proof that finite parity games are determined with memoryless strategies using induction on the number of relevant states. The closest proof that relies on induction over non absorbing states is the one of Graedel [5]. However instead of focusing on a single appropriate vertex for induction as we do here, he considers two reduced games per vertex, for all the vertices of the game. The idea of reasoning about a single state has been inspired to me by the analysis of finite stochastic priority games by Karelovic and Zielonka [6].

7.12. Interval iteration algorithm for MDPs and IMDPs

Markov Decision Processes (MDP) are a widely used model including both non-deterministic and probabilistic choices. Minimal and maximal probabilities to reach a target set of states, with respect to a policy resolving non-determinism, may be computed by several methods including value iteration. This algorithm, easy to implement and efficient in terms of space complexity, iteratively computes the probabilities of paths of increasing length. However, it raises three issues: (1) defining a stopping criterion ensuring a bound on the approximation, (2) analysing the rate of convergence, and (3) specifying an additional procedure to obtain the exact values once a sufficient number of iterations has been performed. The first two issues are still open and, for the third one, an upper bound on the number of iterations has been proposed. Based on a graph analysis and transformation of MDPs, we address these problems. First we introduce an interval iteration algorithm, for which the stopping criterion is straightforward. Then we exhibit its convergence rate. Finally we significantly improve the upper bound on the number of iterations required to get the exact values. We extend our approach to also deal with Interval Markov Decision Processes (IMDP) that can be seen as symbolic representations of MDPs.

7.13. Alignment-Based Trace Clustering

Published in [9].

A novel method to cluster event log traces is presented in this paper. In contrast to the approaches in the literature, the clustering approach of this paper assumes an additional input: a process model that describes the current process. The core idea of the algorithm is to use model traces as centroids of the clusters detected, computed from a generalization of the notion of alignment. This way, model explanations of observed behavior are the driving force to compute the clusters, instead of current model agnostic approaches, e.g., which group log traces merely on their vector-space similarity. We believe alignment-based trace clustering provides results more useful for stakeholders. Moreover, in case of log incompleteness, noisy logs or concept drift, they can be more robust for dealing with highly deviating traces. The technique of this paper can be combined with any clustering technique to provide model explanations to the clusters computed. The proposed technique relies on encoding the individual alignment problems into the (pseudo-)Boolean domain, and has been implemented in our tool DarkSider that uses an open-source solver.

7.14. Aligning Modeled and Observed Behavior: A Compromise Between Complexity and Quality

Published in [17].

Certifying that a process model is aligned with the real process executions is perhaps the most desired feature a process model may have: aligned process models are crucial for organizations, since strategic decisions can be made easier on models instead of on plain data. In spite of its importance, the current algorithmic support for computing alignments is limited: either techniques that explicitly explore the model behavior (which may be worst-case exponential with respect to the model size), or heuristic approaches that cannot guarantee a solution, are the only alternatives. In this paper we propose a solution that sits right in the middle in the complexity spectrum of alignment techniques; it can always guarantee a solution, whose quality depends on the exploration depth used and local decisions taken at each step. We use linear algebraic techniques in combination with an iterative search which focuses on progressing towards a solution. The experiments show a clear reduction in the time required for reaching a solution, without sacrificing significantly the quality of the alignment obtained.

7.15. Temporal Reprogramming of Boolean Networks

Published in [15].

Cellular reprogramming, a technique that opens huge opportunities in modern and regenerative medicine, heavily relies on identifying key genes to perturb. Most of computational methods focus on finding mutations to apply to the initial state in order to control which attractor the cell will reach. However, it has been shown, and is proved in this article, that waiting between the perturbations and using the transient dynamics of the system allow new reprogramming strategies. To identify these temporal perturbations, we consider a qualitative model of regulatory networks, and rely on Petri nets to model their dynamics and the putative perturbations. Our method establishes a complete characterization of temporal perturbations, whether permanent (mutations) or only temporary, to achieve the existential or inevitable reachability of an arbitrary state of the system. We apply a prototype implementation on small models from the literature and show that we are able to derive temporal perturbations to achieve trans-differentiation.

7.16. Goal-Driven Unfolding of Petri Nets

Published in [10].

Unfoldings provide an efficient way to avoid the state-space explosion due to interleavings of concurrent transitions when exploring the runs of a Petri net. The theory of adequate orders allows one to define finite prefixes of unfoldings which contain all the reachable markings. In this paper we are interested in reachability of a single given marking, called the goal. We propose an algorithm for computing a finite prefix of the unfolding of a 1-safe Petri net that preserves all minimal configurations reaching this goal. Our algorithm combines the unfolding technique with on-the-fly model reduction by static analysis aiming at avoiding the exploration of branches which are not needed for reaching the goal. We present some experimental results.

7.17. Euler's Method Applied to the Control of Switched Systems

Hybrid systems are a powerful formalism for modeling and reasoning about cyber-physical systems. They mix the continuous and discrete natures of the evolution of computerized systems. Switched systems are a special kind of hybrid systems, with restricted discrete behaviours: those systems only have finitely many different modes of (continuous) evolution, with isolated switches between modes. Such systems provide a good balance between expressiveness and controllability, and are thus in widespread use in large branches of industry such as power electronics and automotive control. The control law for a switched system defines the way of selecting the modes during the run of the system. Controllability is the problem of (automatically) synthesizing a control law in order to satisfy a desired property, such as safety (maintaining the variables within a given zone) or stabilisation (confinement of the variables in a close neighborhood around an objective point).

In order to compute the control of a switched system, we need to compute the solutions of the differential equations governing the modes. Euler's method is the most basic technique for approximating such solutions. We present here an estimation of the Euler's method local error, using the notion of "one-sided Lispchitz constant" for modes. This yields a general control synthesis approach which can encompass several features such as bounded disturbance and compositionality.

7.18. An Improved Algorithm for the Control Synthesis of Nonlinear Sampled Switched Systems

Published in [6].

A novel algorithm for the control synthesis for nonlinear switched systems is presented in this paper. Based on an existing procedure of state-space bisection and made available for nonlinear systems with the help of guaranteed integration, the algorithm has been improved to be able to consider longer patterns of modes with a better pruning approach. Moreover, the use of guaranteed integration also permits to take bounded perturbations and varying parameters into account. It is particularly interesting for safety critical applications, such as in aeronautical, military or medical fields. The whole approach is entirely guaranteed and the induced controllers are correct-by-design. Some experimentations are performed to show the important gain of the new algorithm.

7.19. Distributed Control Synthesis Using Euler's Method

Published in [22].

In a previous work, we explained how Euler's method for computing approximate solutions of systems of ordinary differential equations can be used to synthesize safety controllers for sampled switched systems. We continue here this line of research by showing how Euler's method can also be used for synthesizing safety controllers in a distributed manner. The global system is seen as an interconnection of two (or more) subsystems where, for each component, the sub-state corresponding to the other component is seen as an "input"; the method exploits (a variant of) the notions of incremental input-to-state stability (δ-ISS) and ISS Lyapunov function. We illustrate this distributed control synthesis method on a building ventilation example.

7.20. Control Synthesis of Nonlinear Sampled Switched Systems using Euler's Method

Published in [21].

In this paper, we propose a symbolic control synthesis method for nonlinear sampled switched systems whose vector fields are one-sided Lipschitz. The main idea is to use an approximate model obtained from the forward Euler method to build a guaranteed control. The benefit of this method is that the error introduced by symbolic modeling is bounded by choosing suitable time and space discretizations. The method is implemented in the interpreted language Octave. Several examples of the literature are performed and the results are compared with results obtained with a previous method based on the Runge-Kutta integration method.

7.21. Metastability-Aware Memory-Efficient Time-to-Digital Converter

Published in [11].

We propose a novel method for transforming delay- line time-to-digital converters (TDCs) into TDCs that output Gray code without relying on synchronizers. We formally prove that the inevitable metastable memory upsets (Marino, TC'81) do not induce an additional time resolution error. Our modified design provides suitable inputs to the recent metastability-containing sorting networks by Lenzen and Medina (ASYNC'16) and Bund et al. (DATE'17). In contrast, employing existing TDCs would require using thermometer code at the TDC output (followed by conversion to Gray code) or resolving metastability inside the TDC. The former is too restrictive w.r.t. the dynamic range of the TDCs, while the latter loses the advantage of enabling (accordingly much faster) computation without having to first resolve metastability.

Our all-digital designs are also of interest in their own right: they support high sample rates and large measuring ranges at nearly optimal bit-width of the output, yet maintain the original delay-line?s time resolution. No previous approach unifies all these properties in a single device.

7.22. Brief Announcement: Lower Bounds for Asymptotic Consensus in Dynamic Networks

Published in [12].

7.23. Metastability Tolerant Computing

Published in [16].

Synchronization using flip-flop chains imposes a latency of a few clock cycles when transferring data and control signals between clock domains. We propose a design scheme that avoids this latency by performing synchronization as part of state/data computations while guaranteeing that metastability is contained and its effects tolerated (with an acceptable failure probability). We present a theoretical framework for modeling synchronous state machines in the presence of metastability and use it to prove properties that guarantee some form of reliability. Specifically, we show that the inevitable state/data corruption resulting from propagating metastable states can be confined to a subset of computations. Applications that can tolerate certain failures can exploit this property to leverage low-latency and quasi-reliable operation simultaneously. We demonstrate the approach by designing a Network-on-Chip router with zero-latency asynchronous ports and show via simulation that it outperforms a variant with two flip-flop synchronizers at a negligible cost in packet transfer reliability.

PARSIFAL Project-Team

6. New Results

6.1. Separating Functional Computation from Relations

Participants: Ulysse Gérard, Dale Miller.

The logical foundation of arithmetic generally starts with a quantificational logic over relations. Of course, one often wishes to have a formal treatment of functions within this setting. Both Hilbert and Church added choice operators (such as the epsilon operator) to logic in order to coerce relations that happen to encode functions into actual functions. Others have extended the term language with confluent term rewriting in order to encode functional computation as rewriting to a normal form (e.g., the Dedukti proof checking project [46]) It is possible to take a different approach that does not extend the underlying logic with either choice principles or with an equality theory. Instead, we use the familiar two-phase construction of focused proofs and capture functional computation entirely within one of these phases. As a result, computation of functions can remain purely relational even when it is computing functions. This result, which appearred in [22], could be used to add to the Abella theorem prover a primitive method for doing deterministic computations.

6.2. Translating between implicit and explicit versions of proof

Participants: Roberto Blanco, Zakaria Chihani, Dale Miller.

As we have demonstrated within the Parsifal team, the Foundational Proof Certificate (FPC) framework can be used to define the semantics of a wide range of proof evidence. We have given such definitions for a number of textbook proof systems as well as for the proof evidence output from some existing theorem proving systems. An important decision in designing a proof certificate format is the choice of how many details are to be placed within certificates. Formats with fewer details are smaller and easier for theorem provers to output but they require more sophistication from checkers since checking will involve some proof reconstruction. Conversely, certificate formats containing many details are larger but are checkable by less sophisticated checkers. Since the FPC framework is based on well-established proof theory principles, proof certificates can be manipulated in meaningful ways. In fact, we have shown how it is possible to automate moving from implicit to explicit (*elaboration*) and from explicit to implicit (*distillation*) proof evidence via the proof checking of a *pair of proof certificates*. Performing elaboration makes it possible to transform a proof certificate with details missing into a certificate packed with enough details so that a simple kernel (without support for proof reconstruction) can check the elaborated certificate. This design allows us to trust in only a single, simple checker of explicitly described proofs but trust in a range of theorem provers employing a range of proof structures. Experimental results of using this design appear in

6.3. Combinatorial Flows

Participant: Lutz Straßburger.

Combinatorial flows are a variation of combinatorial proofs that allow for the substitution of proofs into proofs (instead of just substituting formulas). This makes combinatorial flows p-equvalent to Frege systems with substitution, which are the strongest proof systems with respect to p-simulation, as studied in proof complexity. Since combinatorial flows have a polynomial correctness criterion, they can also be seen as an improvement to atomic flows (which do not have a correctness criterion). This work has been presented at the FCSD 2017 conference [37], [28]

6.4. Justification Logic for Constructive Modal Logic

Participants: Lutz Straßburger, Sonia Marin.

Justification logic is a family of modal logics generalizing the Logic of Proofs LP, introduced by Artemov in [45]. The original motivation, which was inspired by works of Kolmogorov and Gödel in the 1930's, was to give a classical provability semantics to intuitionistic propositional logic. The language of the Logic of Proofs can be seen as a modal language where occurrences of the $\Box$-modality are replaced with terms, also known as *proof polynomials*, *evidence terms*, or *justification terms*, depending on the setting. The intended meaning of the formula '$t : A$' is '*t is a proof of A*' or, more generally, the reason for the validity of A. Thus, the justification language is viewed as a refinement of the modal language, with one provability construct $\Box$ replaced with an infinite family of specific proofs. In a joint work with Roman Kuznets (TU Wien), we add a second type of terms, which we call *witness terms* and denote by Greek letters. Thus, a formula $\Diamond A$ is to be realized by '$\mu : A$'. The intuitive understanding of these terms is based on the view of $\Diamond$ modality as representing consistency (with $\Box$ still read as provability). The term μ justifying the consistency of a formula is viewed as an abstract witnessing model for the formula. We keep these witnesses abstract so as not to rely on any specific semantics. All the operations on witness terms that we employ to ensure the realization theorem for CK, CD, CT, and $CS4$. This work has been presented at the IMLA 2017 workshop [40]

6.5. Proof Theory of Indexed Nested Sequents

Participants: Lutz Straßburger, Sonia Marin.

Indexed nested sequents are an extension of nested sequents allowing a richer underlying graph-structure that goes beyond the plain tree-structure of pure nested sequents. For this reason they can be used to give deductive systems to modal logics which cannot be captured by pure nested sequents. In this work we show how the standard cut-elimination procedure for nested sequents can be extended to indexed nested sequents, and we discuss how indexed nested sequents can be used for intuitionistic modal logics. These results have been presented at the TABLEAUX 2017 conference [24], [35]

6.6. On the Length of Medial-Switch-Mix Derivations

Participant: Lutz Straßburger.

Switch and medial are two inference rules that play a central role in many deep inference proof systems. In specific proof systems, the mix rule may also be present. In a joint work with Paola Bruscoli (University of Bath) we show that the maximal length of a derivation using only the inference rules for switch, medial, and mix, modulo associativity and commutativity of the two binary connectives involved, is quadratic in the size of the formula at the conclusion of the derivation. This shows, at the same time, the termination of the rewrite system. This result has been presented at the International Workshop on Logic, Language, Information, and Computation 2017 [20].

6.7. Maehara-style Modal Nested Calculi

Participant: Lutz Straßburger.

In a joint work with Roman Kuznets (TU Wien), we develop multi-conclusion nested sequent calculi for the fifteen logics of the intuitionistic modal cube between IK and IS5. The proof of cut-free completeness for all logics is provided both syntactically via a Maehara-style translation and semantically by constructing an infinite birelational countermodel from a failed proof search [83]. Interestingly, the Maehara-style translation for proving soundness syntactically fails due to the hierarchical structure of nested sequents. Consequently, we only provide the semantic proof of soundness. The countermodel construction used to prove completeness required a completely novel approach to deal with two independent sources of non-termination in the proof search present in the case of transitive and Euclidean logics.

6.8. Combining inference systems in the CDSAT framework

Participant: Stéphane Graham-Lengrand.

In 2016 we had designed a methodology [49], based on *inference systems*, for combining theories in SMT-solving, that supersedes the existing approaches, namely that of Nelson-Oppen [78] and that of MCSAT [86], [66]. While soundness and completeness of our approach were proved in 2016, we further developed, in 2017, the meta-theory of this system, now called CDSAT for *Conflict-Driven Satisfiability*, in particular with

- a proof of termination for the CDSAT system, and the identification of sufficient conditions, on the theory modules to be combined, for the global termination of the system to hold;
- a learning mechanism, whereby the system discovers lemmas along the run, which can be used later to speed-up the rest of the run;
- an enrichment of the CDSAT system with proof-object generation, and the identification of proof-construction primitives that can be used to make the answers produced by CDSAT correct-by-construction.

The first result, together with the introduction of the CDSAT framework, was publishing this year in [19]. The last two results are described in a paper accepted for publication at CPP in 2018.

6.9. Theory modules for CDSAT

Participant: Stéphane Graham-Lengrand.

The CDSAT system described above is a framework for the combination of theory modules, so it is only useful inasmuch many theories can be captured as CDSAT theory modules. Theory modules are essentially given by a set of inference rules and, for each input problem, a finite set of expressions that are allowed to be used by CDSAT at runtime. These ingredients need to satisfy some requirement for soundness, completess, and termination of CDSAT. In 2017 we identified such theory modules for the following theories

- Boolean logic;
- Linear Rational Arithmetic;
- Equality with Uninterpreted Function symbols;
- Any theory whose ground satisfiability is decidable, if one is willing to give up the fine-grained aspect of inference rules;
- Bitvectors (core fragment).

The first four cases of theories were published in [19], while the Bitvector theory was published in [21].

6.10. Environments and the Complexity of Abstract Machines

Participant: Beniamino Accattoli.

This joint work with Bruno Barras (Inria) [30] belongs to line of work *Cost Models and Abstract Machines for Functional Languages*, supported by the ANR project COCA HOLA.

We study various notions of environments (local, global, split) for abstract machines for functional languages, from a complexity and implementative point of view.

An environment is a data structure used to implement sharing of subterms. There are two main styles. The most common one is to have many local environments, one for every piece of code in the data structures of the machine. A minority of works uses a single global environment instead. Up to now, the two approaches have been considered equivalent, in particular at the level of the complexity of the overhead: they have both been used to obtain bilinear bounds, that is, linear in the number of beta steps and in the size of the initial term.

Our main result is that local environments admit implementations that are asymptotically faster than global environments, lowering the dependency from the size of the initial term from linear to logarithmic, thus improving the bounds in the literature. We also show that a third style, split environments, that are in between local and global ones, has the benefits of both. Finally, we provide a call-by-need machine with split environments for which we prove the new improved bounds on the overhead.

6.11. The Negligible and Yet Subtle Cost of Pattern Matching

Participant: Beniamino Accattoli.

This joint work with Bruno Barras (Inria) [31] belongs to line of work *Cost Models and Abstract Machines for Functional Languages*, supported by the ANR project COCA HOLA.

In this work we extend results about time cost models for the λ-calculus to a larger language, namely the λ-calculus with constructors and pattern matching. We consider all natural evaluation strategies, that is, call-by-name, call-by-value, and call-by-need.

The results are expected, and considered folklore, but we show that the question is subtler than it seems at first sight, by exhibiting some counter-example for naive formulations of the extensions. The, we show the actual results for the right extensions.

6.12. Implementing Open Call-by-Value

This joint work with Giulio Guerrieri (Oxford University) [32] belongs to line of work *Cost Models and Abstract Machines for Functional Languages*, supported by the ANR project COCA HOLA.

The theory of the call-by-value λ-calculus relies on weak evaluation and closed terms, that are natural hypotheses in the study of programming languages. To model proof assistants, however, strong evaluation and open terms are required. Open call-by-value is the intermediate setting of weak evaluation with open terms, on top of which Grégoire and Leroy designed the abstract machine of Coq. This paper provides a theory of abstract machines for open call-by-value. The literature contains machines that are either simple but inefficient, as they have an exponential overhead, or efficient but heavy, as they rely on a labelling of environments and a technical optimization. We introduce a machine that is simple and efficient: it does not use labels and it implements open call-by-value within a bilinear overhead. Moreover, we provide a new fine understanding of how different optimizations impact on the complexity of the overhead.

6.13. Further Formalizing the Meta-Theory of Linear Logic

Participants: Kaustuv Chaudhuri, Leonardo Lima, Giselle Reis.

We have continued our formalization of the meta-theory of substructural logics by giving a fully formal proof of cut-elimination (and hence of completeness) for focused classical first-order linear logic. This is the first time that this complete system has had a fully formalized proof.

This formalization serves as a *tour de force* of Abella's ability to reason about mutual induction and support sophisticated binding constructs.

An extended invited paper is currently under review, to possibly appear in a special issue of *Theoretical Computer Science* in 2018.

6.14. Formalized Meta-Theory of Simultaneous Substitutions

Participant: Kaustuv Chaudhuri.

It has long been claimed that a logical framework must have sophisticated built-in support for reasoning about formal substitutions in order to formalize relational meta-theorems such as strong normalization (using a logical relations style argument) or that applicative simulation is a pre-congruence. A number of type-theoretic frameworks in recent years, such as Beluga, have indeed started to incorporate such constructs in their core systems.

We have recently shown how to implement the meta-theory of simultaneous substitutions in the Abella system without any modification or extension of the (trusted) kernel, and without sacrificing any expressivity. The results of this paper will appear in the ACM Conference on Certified Programming in Jannuary 2018.

Our hope is that this work will be continued in the near future to build a specification language based on contextual LF in Abella, similar to how Abella/LF handles (ordinary) LF.

6.15. Hybrid Linear Logic Revisited

Participants: Kaustuv Chaudhuri, Joëlle Despeyroux, Carlos Olarte, Elaine Pimentel.

We have written a comprehensive account of hybrid linear logic (HyLL) and its relation to a number of related linear logic variants such as subexponential logic. One of the new and novel examples that we have fully worked out is how to encode CTL and CTL* in HyLL, which shows that HyLL can indeed serve as a logical framework for representing and reasoning about constrained transition systems, such as biochemical networks.

This account will appear in a special issue of MSCS in 2018.

6.16. Correctness of Speculative Optimizations with Dynamic Deoptimization

Participant: Gabriel Scherer.

This joint work with Olivier Flückiger, Ming-Ho Yee Ming-Ho, Aviral Goel, Amal Ahmed and Jan Vitek was initiated during Gabriel Scherer's post-doctoral stay at Northeastern University, Boston, USA.

Practitioners from the software industry find it difficult to implement Just-In-Time (JIT) compilers for dynamic programming languages, such as Javascript: they don't know how to reason on the correctness of their optimizations in the context of Just-In-Time code generation and deoptimization. We explain how to adapt reasoning approaches and proof techniques from standard compiler research to this new setting.

This work [14] will appear in POPL 2018.

PI.R2 Project-Team

6. New Results

6.1. Effects in proof theory and programming

Participants: Hugo Herbelin, Étienne Miquey, Yann Régis-Gianas, Alexis Saurin.

6.1.1. A classical sequent calculus with dependent types

Dependent types are a key feature of type systems, typically used in the context of both richly-typed programming languages and proof assistants. Control operators, which are connected with classical logic along the proof-as-program correspondence, are known to misbehave in the presence of dependent types [14], unless dependencies are restricted to values. As a step in his work to develop a sequent-calculus version of Hugo Herbelin's dPA_ω system [16], Étienne Miquey proposed a sequent calculus with classical logic and dependent types. His calculus—named dL—is an extension of the $\mu\tilde{\mu}$-calculus with a syntactical restriction of dependent types to the fragment of *negative-elimination free* proofs. The corresponding type system includes a list of explicit dependencies, which maintains type safety. He showed that a continuation-passing style translation can be derived by adding delimited continuations, and how a chain of dependencies can be related to a manipulation of the return type of these continuations. This work has been presented at ESOP 2017 [39].

6.1.2. Normalisation and realisability interpretation of call-by-need with control

The call-by-need evaluation strategy is an evaluation strategy of the λ-calculus which evaluates arguments of functions only when needed, and, when needed, shares their evaluations across all places where the argument is needed. The call- by-need evaluation is for instance at the heart of a functional programming language such as Haskell. A continuation-passing-style semantics for call-by-need, de facto giving a semantics to control operators, was proposed in the 90s by Okasaki, Lee and Tarditi. However, this semantics does not ensure normalisation of simply-typed call-by-need evaluation, thus failing to ensure a property which holds in the simply-typed call-by-name and call-by-value cases. Étienne Miquey and Hugo Herbelin have been considering a call-by-need λ-calculus due to Ariola et al. for which they proved the normalisation by means of a realisability interpretation. Incidentally, the variant of realisability they proposed allows to define realisers as pairs of a term and a substitution. This paves the way to give interpretation of calculus with global and mutable memory. This work has been accepted for publication at the FOSSACS 2018 conference.

6.1.3. A sequent calculus with dependent types for classical arithmetic

In 2012, Hugo Herbelin showed that classical arithmetic in finite types extended with strong elimination of existential quantification proves the axiom of dependent choice. Getting classical logic and choice together without being inconsistent is made possible by: (1) constraining strong elimination of existential quantification to proofs that are essentially intuitionistic; (2) turning countable universal quantification into an infinite conjunction of classical proofs, which are evaluated along a call-by-need evaluation strategy, so as to extract from them intuitionistic contents that complies to the intuitionistic constraint put on strong elimination of existential quantification.

Relying on its sequent calculus with dependent types and its realisability interpretation for call-by-need with control, Étienne Miquey proposed in his thesis a sequent calculus with the same computational features [24]. His calculus therefore also allows for the direct definition of proof terms for the axioms of countable and dependent choices. The proofs of normalisation and soundness are made through a realisability interpretation of the calculus, which is obtained by using Danvy's methodology of semantics artifacts.

6.1.4. Reverse mathematics of Gödel's completeness theorem

Charlotte Barot, under the supervision of Hugo Herbelin, studied the relative intuitionistic strength of Gödel's completeness theorem, the ultrafilter lemma, and different forms of the Fan Theorem, as a way to transfer computational contents of proofs from one to the other theorems.

6.1.5. A theory of effects and resources

Arnaud Spiwack, in collaboration with Jean-Philippe Bernardy, Mathieu Boespflug, Ryan R. Newton and Simon Peyton-Jones, developed an extension of the type system of Haskell with linear types The work is to be presented at POPL'18.

In collaboration with Thomas Letan (Agence Nationale pour la Sécurité des Systèmes Informatiques), Yann Régis-Gianas studied how free monads can be used to develop modular implementations and proofs of effectful systems. This proof technique is applied to the formal study of architectural attacks on IBM PC like architectures.

6.1.6. Classical realisability and implicative algebras

Étienne Miquey has been working with Alexandre Miquel in Montevideo on the topic of implicative algebras. Implicative algebras are an algebraisation of the structure needed to develop a realisability model. In particular, they give rise to the usual ordered combinatory algebras and thus to the triposes used to model classical realisability. An implicative algebra is given by an implicative structure (which consists of a complete semi-lattice with a binary operation $\rightarrow$) together with a separator containing the element interpreted as true in the structure. Following the work of Guillaume Munch-Maccagnoni on focalisation and classical realisability, Étienne Miquey gave alternative presentations within structures based on other connectives rather than $\rightarrow$, namely disjunctive algebras (based on negation, "par") and conjunctive algebras (negation, tensor). Such connectives correspond to the decomposition of the arrow according to the strategy of evaluation (call-by-name/call-by-value). In particular, he showed that disjunctive algebras were particular cases of implicative algebras; and that conjunctive algebras can be obtained by duality from disjunctive algebras. Besides, Étienne Miquey has formalised the theory of implicative algebras (resp. disjunctive, conjunctive) in Coq.

6.2. Reasoning and programming with infinite data

Participants: Amina Doumane, Yann Régis-Gianas, Alexis Saurin.

This theme is part of the ANR project Rapido (see the National Initiatives section).

6.2.1. Proof theory of infinitary and circular proofs

In collaboration with David Baelde and Guilhem Jaber, Amina Doumane and Alexis Saurin extended the proof theory of infinite proofs for fixpoint logics by relaxing the validity condition necessary to distinguish sound proofs from invalid ones. In CSL 2016, Baelde, Doumane and Saurin proved cut-elimination and focalisation for infinite proofs for $\mu MALL$ with a validity condition inspired from the acceptance condition of parity automata (or the winning condition of parity games). However, this validity condition rules out lots of proofs which are computational sound and does not account for the cut-axiom interaction in sequent proofs.

With Jaber, they relaxed the validity condition to allow infinite branches to be supported by threads bouncing on axioms and cuts. This allows for a much more flexible criterion, inspired from Girard's geometry of interaction, approximating productivity. If the decidability of the validity condition in the most general case is still open, it allows for decidable restrictions which are still useful in the sense they allow for a much more flexible writing of circular proofs (or, through the proofs-as-programs bridge, circular progams). Cut-elimination is obtained in two steps, combining CSL 2016 result with a technique for "straightening" bouncing threads, that is perfoming just the necessary amount of cut-elimination to recover straight threads, the two results are combined thanks to a compression lemma, a standard result from infinitary rewriting ensuring that a transfinite strongly converging sequence can be turned into an ω-indexed strongly converging sequence. Preliminary results were presented at the Types 2017 conference.

6.2.2. Automata theory meets proof theory: completeness of the linear time mu-calculus.

Amina Doumane extended her previous results with David Baelde, Lucca Hirschi and Alexis Saurin proving a constructive completeness theorem for the full linear-time μ-calculus, while the previous results only captured a fragment of the linear-time mu-calculus expressing all inclusions of Büchi automata suitably encoded as formulas.

In order to achieve this tour de force (for which her publication at LICS 2017 received the Kleene award of the best student paper [37], see Highlights of the year), she identified several fragments of the linear-time mu-calculus corresponding to various classes of ω-automata and proved completeness of those classes by using circular proof systems and finitisation of the infinite proofs in the Kozen's usual axiomatisation (see paragraph on finitising circular proofs for more details).

6.2.3. Brotherston-Simpson's conjecture: Finitising circular proofs

An important and most active research topic on circular proofs is the comparison of circular proof systems with usual proof systems with induction and co-induction rules à la Park. This can be viewed as comparing the proof-theoretical power of usual induction reasoning with that of Fermat's infinite descent method. Berardi and Tatsuta, as well as Simpson, obtained in 2017 important results in this direction for logics with inductive predicates à la Martin-Löf. Those frameworks, however, are weaker than those of fixpoint logic which can express and mix least and greatest fixpoints by interleaving μ and ν statements.

In the setting of fixpoint logics with circular proofs, several investigations were carried on in the team:

- firstly, in the setting of the usual validity condition for circular proofs of $\mu MALL$, Doumane extended in her PhD thesis a translatibility criterion for finitising circular proofs which was first used in joint work with Baelde, Saurin and Hirschi and later applied to the full linear-time mu-calculus in her LICS 2017 paper. Her translatibility criterion abstracts the proof scheme for finitising circular proofs and is not formulated with respect to a specific fragment of the logic, but with respect to conditions allowing finitisation of the cycles.
- Secondly, Nollet, working with Saurin and Tasson, recently proposed a new validity condition which is quite straightfoward to check (it can be checked at the level of elementary cycles of the circular proofs, while the other criteria need to check a condition on every infinite branch) and still capture all circular proofs obtained from $\mu MALL$ finite proofs. The condition for cycling in those proofs is more constrained than that of Baelde, Doumane and Saurin but the proof contains more information which can be used to exctract inductive invariants. With this validity condition which can be useful for proof search for circular proofs, they obtained partial finitisation results and are currently aiming at solving the most general Brotherston-Simpson's conjecture.

6.2.4. Co-patterns

In collaboration with Paul Laforgue (Master 2, University Paris 7), Yann Régis-Gianas developed an extension of OCaml with copatterns. Copatterns generalize standard ML patterns for algebraic datatypes: While a pattern-matching destructs a finite value defined using a constructor, a copattern-matching creates an infinite computation defined in terms of its answers to observations performed by the evaluation context. They exploits the duality between functions defined by pattern matching and functions that define codata by copattern-matching, going from the second to the first by introducing a well-typed inversion of control which is a purely local syntactic transformation. This result shows that copattern-matching can be added with no effort to any programming language equipped with second-order polymorphism and generalized algebraic datatypes. This work has been published in the proceeding of PPDP'17. A short paper has also been accepted at JFLA'18.

6.2.5. Streams, classical logic and the ordinal λ-calculus

Polonsky and Saurin defined an extension of infinitary λ-calculi allowing transfinite iteration of abstraction and ordinal sequences of applications, Λ^o, and established a standardisation theorem for this calculus. The $\Lambda\mu$-calculus can be embedded in this calculus, as well as Saurin's full Stream hierarchy: as a consequence, they obtain a uniform framework to investigate this family of calculi and provide uniform proofs of important results such a standardisation.

6.2.6. Theory of fixpoints in the lambda-calculus

In collaboration with Manzonetto, Polonsky and Simonsen, Saurin studied two long-standing conjectures on fixpoints in the λ-calculus: the "fixpoint property" and the "double-fixpoint conjecture". The former asserts that every λ-term admits either a unique or an infinite number of β-distinct fixpoints while the second,

formulated by Statman, says that there is no fixpoint satisfying $Y\delta = Y$ for $\delta = \lambda y, x.x(yx)$. They proved the first conjecture in the case of open terms and refute it in the case of sensible theories (instead of β). Moreover, they provide sufficient conditions for both conjectures in the general case. Concerning the double-fixpoint conjecture, they propose a proof technique identifying two key properties from which the results would follow, while they leave as conjecture to prove that those actually hold. Those results are currently submitted to a journal [53].

6.3. Effective higher-dimensional algebra

Participants: Cyrille Chenavier, Pierre-Louis Curien, Yves Guiraud, Cédric Ho Thanh, Maxime Lucas, Philippe Malbos, Samuel Mimram, Jovana Obradović, Matthieu Sozeau.

6.3.1. Higher linear rewriting

Yves Guiraud and Philippe Malbos have completed a four-year long collaboration with Eric Hoffbeck (LAGA, Univ. Paris 13), whose aim was to develop a theory of rewriting in associative algebras, with a view towards applications in homological algebra. They adapted the known notion of polygraph [69] to higher-dimensional associative algebras, and used these objects to develop a rewriting theory on associative algebras that generalises the two major tools for computations in algebras: Gröbner bases [68] and Poincaré-Birkhoff-Witth bases [105]. Then, they transposed the construction of [12], based on an extension of Squier's theorem [108] in higher dimensions, to compute small polygraphic resolutions of associative algebras from convergent presentations. Finally, this construction has been related to the Koszul homological property, yielding necessary or sufficient conditions for an algebra to be Koszul. The resulting work has just been submitted for publication [47].

Cyrille Chenavier has continued his work on reduction operators, a functional point of view on rewriting in associative algebras initiated by Berger [62], on which his PhD thesis was focused [4]. First, using the lattice structure of the reduction operators, he gave a new algebraic characterisation of confluence, and developed a new algorithm for completion, based on an iterated use of the meet-operation of the lattice [28]. Then he related this completion procedure to Faugère's F4 completion procedure for noncommutative Gröbner bases [79]. Finally, he gave a construction of a linear basis of the space of syzygies of a set of reduction operations, and used this work to optimise his completion procedure [45].

6.3.2. Cubical higher algebra

Maxime Lucas, supervised by Yves Guiraud and Pierre-Louis Curien, has applied the rewriting techniques of Guiraud and Malbos [92] to prove coherence theorems for bicategories and pseudofunctors. He obtained a coherence theorem for pseudonatural transformations thanks to a new theoretical result, improving on the former techniques, that relates the properties of rewriting in 1- and 2-categories [31]. Then he has transposed to a cubical setting, and improved, the results of [12]. This first involved a deep foundational work on the connections between globular and cubical higher categories [51], generalising several already known links in a unique theoretical setting [66], [67], [57], [110]. Then, he could prove Squier's theorem, giving a construction of a polygraphic resolution of monoids in the category of cubical Gray monoids [50]. All these results are contained in his PhD thesis, that was successfully defended in December 2017 [23].

6.3.3. Coherent Presentations of Monoidal categories

Presentations of categories are a well-known algebraic tool to provide descriptions of categories by means of generators, for objects and morphisms, and relations on morphisms. Pierre-Louis Curien and Samuel Mimram have generalised this notion, in order to consider situations where the objects are considered modulo an equivalence relation, which is described by equational generators. When those form a convergent (abstract) rewriting system on objects, there are three very natural constructions that can be used to define the category which is described by the presentation: one consists in turning equational generators into identities (i.e. considering a quotient category), one consists in formally adding inverses to equational generators (i.e. localising the category), and one consists in restricting to objects which are normal forms. Under suitable coherence conditions on the presentation, the three constructions coincide, thus generalising celebrated results on presentations of groups. Those conditions are then extended to presentations of monoidal categories [29].

6.3.4. Categorified cyclic operads

The work of Pierre-Louis Curien and Jovana Obradović on categorified cyclic operads has been condtionally accepted in the Journal Applied Categorical Structures [46]. The revision will include a careful treatment of weakened identity laws, as well of weakened equivariance laws. It will also include the details of an example and an illustration of the work. The example involves a generalisation of profunctors, and the application is to the notion of anti-cyclic operad, which they prove to be "sign-coherent".

6.3.5. Syntactic aspects of hypergraph polytopes

In collaboration with Jelena Ivanović, Pierre-Louis Curien and Jovana Obradović have introduced an inductively defined tree notation for all the faces of polytopes arising from a simplex by truncations, that allows them to view inclusion of faces as the process of contracting tree edges. This notation instantiates to the well-known notations for the faces of associahedra and permutohedra. Various authors have independently introduced combinatorial tools for describing such polytopes. In this work, the authors build on the particular approach developed by Došen and Petrić, who used the formalism of hypergraphs to describe the interval of polytopes from the simplex to the permutohedron. This interval was further stretched by Petrić to allow truncations of faces that are themselves obtained by truncations, and iteratively so. The notation applies to all these polytopes, and this fact is illustrated by showing that it instantiates to a notation for the faces of the permutohedron-based associahedra, that consists of parenthesised words with holes. In their work, Pierre-Louis Curien, Jovana Obradović and Jelena Ivanović also explore links between polytopes and categorified operads, as a follow-up of another work of Došen and Petrić, who had exhibited some families of hypergraph polytopes (associahedra, permutohedra, and hemiassociahedra) describing the coherences, and the coherences between coherences etc., arising by weakening sequential and parallel associativity of operadic composition. Their work is complemented with a criterion allowing to recover the information whether edges of these "operadic polytopes" come from sequential, or from parallel associativity. Alternative proofs for some of the original results of Došen and Petrić are also given. A paper containing this material has been accepted in the Journal Homotopy and Related Structure [32].

6.3.6. Opetopes

Opetopes are a formalisation of higher many-to-one operations leading to one of the approaches for defining weak ω-categories. Opetopes were originally defined by Baez and Dolan. A reformulation (leading to a more carefully crafted definition) has been later provided by Batanin, Joyal, Kock and Mascari, based on the notion of polynomial functor. Pierre-Louis Curien has developped a corresponding syntax, which he presented at the workshop "Categories for homotopy and rewriting" (CIRM, September 2017).

Cédric Ho Thanh started his PhD work around opetopes in September 2017. His first contributions include a careful embedding of opetopic sets into polygraphs, and a (finite) critical pair lemma for opetopic sets. Indeed, opetopic sets seem to delimit a subset of polygraphs in which the basics of rewriting theory can be developped, without the anomalies already observed by Lafont and others happening, like the existence of a possibly infinite set of critical pairs in a rewriting system specified by finitely many rules. Opetopes are tree-like and hence first-order-term-like and that is the intuitive reason why these anomalies are avoided.

6.3.7. Higher Garside theory

Building on [9], Yves Guiraud is currently finishing with Matthieu Picantin (IRIF, Univ. Paris 7) a work that generalises already known constructions such as the bar resolution, several resolutions defined by Dehornoy and Lafont [77], and the main results of Gaussent, Guiraud and Malbos on coherent presentations of Artin monoids [10], to monoids with a Garside family. This allows an extension of the field of application of the rewriting methods to other geometrically interesting classes of monoids, such as the dual braid monoids.

Still with Matthieu Picantin, Yves Guiraud develops an improvement of the classical Knuth-Bendix completion procedure, called the KGB (for Knuth-Bendix-Garside) completion procedure. The original algorithm tries to compute, from an arbitrary terminating rewriting system, a finite convergent presentation by adding relations to solve confluence issues. Unfortunately, this algorithm fails on standard examples, like most Artin monoids with their usual presentations. The KGB procedure uses the theory of Tietze transformations, together

with Garside theory, to also add new generators to the presentation, trying to reach the convergent Garside presentation identified in [9]. The KGB completion procedure is partially implemented in the prototype Rewr, developed by Yves Guiraud and Samuel Mimram.

6.3.8. Foundations and formalisation of higher algebra

Yves Guiraud has started a collaboration with Marcelo Fiore (Univ. Cambridge) on the foundations of higher-dimensional categories, with the aim to define a general notion of polygraphs for various notions of algebraic structures. This is based on seeing higher categories as n-oids in a specific n-oidal category (a category with n monoidal structures with exchange morphisms between them). With that point of view, a good notion of polygraph can be iteratively defined for monoids in any monoidal category with pullbacks, which is a sufficiently general setting for most purposes.

Eric Finster, Yves Guiraud and Matthieu Sozeau have started to explore the links between combinatorial higher algebra and homotopy type theory, two domains that describe computations with a homotopical point of view. Their first goal is to formalise the rewriting methods of [12] and [10] in homotopy type theory, establishing a first deep connection between the two fields. This direction will be explored further by Antoine Allioux, a PhD student co-directed by Guiraud and Sozeau, starting in February 2018.

6.4. Incrementality

Participants: Thibaut Girka, Yann Régis-Gianas, Kostia Chardonnet.

In collaboration with Colin Gonzalez, Yann Régis-Gianas developed BLACS, a programming framework that applies differential functional programming techniques to the implementation of asynchronous spreadsheets for big data.

In collaboration with Lelio Brun (ENS), Yann Régis-Gianas developed DeltaCoq, a library for certified incremental functional programming. A paper is in preparation.

6.4.1. Incrementality in proof languages

In collaboration with Paolo Giarrusso, Philipp Shuster and Yufei Cai (Univ Marburg, Allemagne), Yann Régis-Gianas developed a new method to incrementalise higher-order programs using formal derivatives and static caching. Yann Régis-Gianas has developed a mechanised proof for this transformation as well as a prototype language featuring efficient derivatives for functional programs. A paper will be submitted to ICFP 2018.

6.4.2. Difference languages

In collaboration with David Mentré (Mitsubishi), Thibaut Girka and Yann Régis-Gianas developed a metatheoretical framework to develop verifiable difference languages in Coq. Such formal differences capture semantic differences between close programs. This work appeared in PPDP'17 [38].

Kostia Chardonnet and Yann Régis-Gianas started the formalisation of difference languages for Java using the framework developed by Thibaut Girka. In particular, Kostia Chardonnet implemented a mechanised small step operational semantics for a large subset of Java. A paper is in preparation.

6.5. Metatheory and development of Coq

Participants: Hugo Herbelin, Pierre Letouzey, Yann Régis-Gianas, Matthieu Sozeau, Cyprien Mangin, Théo Zimmermann.

6.5.1. Homotopy type theory

Hugo Herbelin worked on the computational contents of extensional equality in type theory. Exploiting the idea introduced in Cohen, Coquand, Huber and Mörtberg's Cubical Type Theory of equality as abstraction over a geometrical dimension, he developed a direct-style system of notations for a scoped iterated parametricity semantics. The resulting logic respects equivalence of types by construction, thus providing a simple computational content to the key axiom of Homotopy Type Theory, namely the axiom of univalence.

6.5.2. Proof irrelevance and Homotopy Type Theory

Gaëtan Gilbert (PhD student of N. Tabareau, Gallinette and M. Sozeau, started in 2016) is studying the integration of a new notion of propositions, called *strict* propositions, in the calculus of inductive constructions. This new sort dubbed sProp supports definitional proof-irrelevance (two proofs of a strict proposition are always convertible), while maintaining compatibility with Univalence or Uniqueness of Identity Proofs. The goal of this work is to provide a more comfortable programming experience in the system by allowing more proofs to be identified "for free" during conversion. This should have an impact both on programming with dependent types (avoiding issues with coercions during typechecking) and for the development of homotopy type theory (avoiding "trivial" transports of equality proofs on natural numbers for example). Gaëtan Gilbert has developed a prototype version integrating this extension in Coq.

6.5.3. Extensionality and Intensionality in Type Theory

Théo Winterhalter (internship co-advised by Matthieu Sozeau and Andrej Bauer in 2017, now PhD student at Inria Nantes, co-advised by Nicolas Tabareau and Matthieu Sozeau) studied a translation from extensional to intensional type theory during his internship with Matthieu Sozeau and a general framework for formalising variants of type theory previously with Andrej Bauer at the University of Ljubljana in Slovenia. They developed a revised version of the translation by Nicolas Oury which doesn't require the use of John Major equality nor suspicious axioms associated to it. It results in a mixed translation that can transport derivations of extensional type theory into intensional 2-level type theory (with an original, syntactic presentation of the latter). This allows in principle to use the convenience of the reflection rule of equality in proofs while being able to derive decorated terms checkable by the kernel of a 2-level variant of Coq: one where both a univalent equality and a strict equality with uniqueness of identity proofs can cohabit. They are working on a Coq formalisation of this result using the Template-Coq framework, which will be extracted to a translation plugin to provide this facility in Coq itself.

6.5.4. Dependent pattern-matching

Cyprien Mangin developed a new simplification engine on top of the Equations plugin. This simplification engine is similar to the one of Cockx [72], allowing an interpretation of dependent pattern-matching that is independent of axioms like UIP or Univalence. While refining the implementation, he also designed a few optimisations allowing for a smarter compilation scheme, in terms of the required properties of the objects and the size of the generated proofs. Matthieu Sozeau concentrated on making the treatment of recursive functions more robust and complete, leading to the first tool of this kind for Coq that can handle both mutual and nested structurally recursive functions along with nested well-founded definitions. The elimination principle generation part of the system was adapted accordingly, putting the tool in good position to replace the previous Function tool of Coq that supports neither dependent pattern-matching nor nested fixpoints. Matthieu Sozeau developed a number of examples showcasing the tool, the largest one having actually been first developed by a student of the MPRI 2.7.2 course. An article presenting this tool and the smart case analysis method is in revision [52]. Version 1.0 of the system was released in December 2017. Cyprien Mangin gave a demo / presentation of the tool at the EUTypes Type Theory Tools workshop in January 2017 and will present a poster and demonstration of the new version at PEPM 2018 in Los Angeles.

Thierry Martinez started the implementation of a dependent pattern-matching compilation algorithm in Coq based on the PhD thesis work of Pierre Boutillier and on the internship work of Meven Bertrand. The algorithm based on small inversion and generalisation is the object of a paper to be submitted to the TYPES post-proceedings.

6.5.5. Transferring theorems along isomorphisms

Following his work on theorem transfer along (iso)morphisms, Théo Zimmermann has started to explore more fundamental aspects that are connected to it: the concept of logical relation, which was originally invented to prove behavioral equivalence of programs and served to formalise parametricity, seems, following Hermida, Reddy and Robinson, to correspond to a very generic relational notion of morphism that was precisely the one needed for transfer lemmas.

6.5.6. Unification

Matthieu Sozeau has developped a complete reimplementation of the basic tactics of Coq in terms of the type-inference unification algorithm of Coq. This work is scheduled to be integrated in part in the 8.8 version of Coq due next year. It should provide a clean slate for development of the 9 series of Coq relying solely on an algorithm close to the one studied with Beta Ziliani in [22].

6.5.7. Cumulativity for Inductive Types

Together with Amin Timany (PhD student of Bart Jacobs at KU Leuven), Matthieu Sozeau developed an extension of the Calculus of Inductive Constructions featuring cumulativity for inductive types [43]. This extension is useful for developments using universe polymorphism like Category Theory formalisations and the HoTT library [35] but also crucial to develop syntactic program translations that add structures to types, as advocated by Boulier et al [65], requiring to validate the cumulativity rule on sigma types. They showed the relative consistency of this extension of the calculus using a set-theoretic model, inspired by the one of Lee and Werner [101] for proof-irrelevance. This extension is integrated in the 8.7 release of Coq and involved a large amount of design and implementation work in particular in relation with the unification strategy used in presence of subtyping and delta reduction, extending the framework studied in [33]. An article describing this work is in revision.

6.6. Formalisation work

Participants: Jean-Jacques Lévy, Daniel de Rauglaudre.

6.6.1. Proofs of algorithms on graphs

Jean-Jacques Lévy and Chen Ran (a PhD student of the Institute of Software, Beijing, visiting the Toccata team 9 months until April 2017) pursue their work about formal proofs of algorithms. Their goal is to provide proofs of algorithms which ought to be both checked by computer and easily human readable. If these kinds of proofs exist for algorithms on inductive structures or recursive algorithms on arrays, they seem less easy to design for combinatorial structures such as graphs. In 2016, they completed proofs for algorithms computing the strongly connected components in graphs. There are mainly two algorithms: one by Kosaraju (1978) working in two phases (some formal proofs of it have already been achieved by Pottier with Coq and by Théry and Gonthier with Coq-SSReflect), one by Tarjan (1972) working in a single pass.

Their proofs use a first-order logic with definitions of inductive predicates. This logic is the one defined in the Why3 system (research-team Toccata, Saclay). They widely use automatic provers interfaced with Why3. A minor part of these proofs is also achieved in Coq. The difficulty of this approach is to combine automatic provers and intuitive design.

In 2017, the same proofs were fully completed in Coq-ssreflect by Cohen and Théry, and in Isabelle-HOL by Merz, both proofs with the assistance of J.-J. Lévy. A Fstar proof is also under development. These proofs are between a factor 4 to 8 in length with respect to the initial Why3 proofs, but more importantly they look less human readable, mainly because of the absence of automatic deduction and several technicalities about termination.

Part of this work (Tarjan 1972) was presented at JFLA 2017 in Gourette [40]. A more comprehensive version was presented at the VSTTE 2017 conference in Heidelberg [36]. Scripts of proofs can be found at http://jeanjacqueslevy.net/why3.

6.6.2. Banach-Tarski Paradox

Banach-Tarski Paradox states that, if we admit the axiom of choice, a sphere is equidecomposable into two spheres identical to the initial one. The equidecomposability is a property of geometric objects: two objects (sets) are equidecomposable if they can be partitioned into a same finite number of sets, and each set of the first object is mapped to a set of the second object by only rotations and translations. In other words, one breaks the first object into a finite number of pieces, and with them, one reconstructs the second object. Its pen and paper proof was done in 1924 by Banach and Tarski.

The formal proof was completed this year by Daniel de Rauglaudre, after 9 months, with a result of about 10000 lines of Coq. A paper about it was published in JFR (Journal of Formalized Reasoning) [34].

6.6.3. Univalence for Free

Together with E. Tanter at Inria Chile and N. Tabareau at Inria Nantes, Matthieu Sozeau developed the theory and implementation of an ad-hoc version of univalence. This axiom at the basis of Homotopy Type Theory morally says that all constructions of type theory are invariant under equivalence, which for programming purposes means invariance by isomorphism. Using a carefuly designed variant of the parametriciy translation for type theory, they can show that indeed all type constructors of type theory, except indexed inductive types with non-hset indices respect univalence. In practice, this leads to a type-class based framework for constructing the proofs that values of a given type do indeed transport equivalences/isomorphisms correctly, relying on univalence itself only for universes and in well-delimited places. An article about this work is in revision [56].

6.6.4. Certified compilation and meta-programming

Matthieu Sozeau participates to the CertiCoq project (https://www.cs.princeton.edu/~appel/certicoq) whose aim is to verify a compiler for the Coq programming language down to CompCert C-light which provides itself a certified compilation path to assembly language. The compiler can already be run and most phases are proven correct. As part of this work, Matthieu Sozeau took the lead of the Template-Coq library development originally developed by Gregory Malecha and extended it. Template-Coq provides quoting and unquoting facilities for Coq's kernel syntax and environment to Coq, allowing to reason on the actual definitions checked by the Coq system in Coq itself. For CertiCoq, the quoted type of Coq terms corresponds to its frontend language. The plugin can however be used in many other ways, notably to implement certified syntactic translations from Coq (or extended theories) to Coq, and to develop plugins to the Coq system in Coq itself. Together with Nicolas Tabareau and Simon Boulier in Nantes and Abhishek Anand at Cornell University, they are developing a general plugin for certified meta-programming in the system. It will be presented at CoqPL'18 [41]. Matthieu Sozeau worked in particular on reimplementing the basic typing and conversion algorithms of Coq inside Coq itself, providing a mechanised specification of the implementation of the system that can be used to verify arbitrarily large parts of it. The type inference algorithm developed there is also useful to help writting program translations on the "forgetful" kernel syntax.

SUMO Project-Team

7. New Results

7.1. Analysis and Verification of Quantitative Systems

7.1.1. Diagnosability

Participants : Hugo Bazille, Éric Fabre, Blaise Genest, Loïc Hélouët, Hervé Marchand, Engel Lefaucheux

7.1.1.1. Diagnosability of repairable faults.

Diagnosability (i.e., the existence of a diagnoser detecting faults in partially-observable systems) can be decided in polynomial time, relying on the so-called twin-machine construction. We have examined the case of repairable faults, and a notion of diagnosability that requires the detection of the fault before it is repaired. We have extended a contribution of 2016 to show that diagnosability of faults and of their repair could help counting the number of occurred faults. It was proved [51] that diagnosability with repair is a *PSPACE*-complete problem. We have completed this result, showing that the close notion of P-diagnosability (diagnosability of a fault even after it is repaired) is also *PSPACE*-complete [20].

7.1.1.2. Diagnosability degree of stochastic systems.

For stochastic systems, several diagnosability properties have been defined. The simplest one, also called A-diagnosability, characterizes the fact that after each fault, detection will almost surely occur. We have considered quantitative versions of the problem, to determine how much a system is diagnosable (when it is not diagnosable for sure). This amounts to characterizing the probability that a faulty run will lead to detection. We have proposed several notions of dignosability degree. Their derivation is generally *NP*-hard, but we have identified situations where complexity becomes polynomial. Besides, we have developed techniques to compute the different moments of the detection delay (mean, variance and upper moments). This allows one to compare systems with similar detection degrees, but that can react faster to faults. In some cases, one may be able to tune a system and trade diagnosability degree againts a faster detection. This approach also yields the distribution of fault location (in time) once detection takes place. Given the first moments of the detection delay, one is also able to compute (sometimes tight) bounds on the response time, for example to lower bound the probability that detection takes place at most T seconds/events after the fault [31].

7.1.1.3. The cost of diagnosis.

We addressed diagnosability and its cost for safe Petri nets. In [37] we have defined an energy-like cost model for Petri nets: transitions can consume or restore energy of the system. We then have defined a partial-order representation for state estimation, and extend the cost model and the capacities of diagnosers. Diagnosers are allowed to use additional energy to refine their estimations. Diagnosability is then seen as an energy game: checking whether disambiguation mechanisms are sufficient to allow diagnosability is in *2EXPTIME*, and one can also decide in *2EXPTIME* whether diagnosability under budget constraint holds.

7.1.2. Analysis of timed systems

Participants : Nicolas Markey, Loïc Hélouët

7.1.2.1. Determinizing timed automata.

In [35], we introduce a new formalism called *automata over a timed domain*, which generalizes timed automata; this formalism provides an adequate framework for determinization. In our formalism, determinization w.r.t. timed language is always possible at the cost of changing the timed domain. We give a condition for determinizability of automata over a timed domain *without changing the timed domain*, which allows us to recover several known determinizable classes of timed systems, such as strongly-non-zeno timed automata, integer-reset timed automata, perturbed timed automata, etc. Moreover, in the case of timed automata, this condition encompasses most determinizability conditions from the literature. Our aim now is to extend this work towards more efficient algorithms for monitoring timed systems.

7.1.2.2. Concurrent Timed Systems.

Time Petri nets (TPNs) are a classical extension of Petri nets with timing constraints attached to transitions, for which most verification problems are undecidable. We consider TPNs under a strong semantics with multiple enablings of transitions. This year, we have extened a work started in 2016, focusing on a structural subclass of unbounded TPNs, where the underlying untimed net is free choice, and showed that it enjoys nice properties in the timed setting under a multi-server semantics [46], [25]. In particular, we have showed that the questions of firability (whether a chosen transition can fire), and termination (whether the net has a non-terminating run) are decidable for this class. Next, we have considered the problem of robustness under guard enlargement and guard shrinking, i.e., whether a given property is preserved even if the system is implemented on an architecture with imprecise time measurement. For unbounded free choice TPNs with a multi-server semantics, we have show decidability of robustness of firability and of termination under both guard enlargement and shrinking.

7.2. Control of Quantitative Systems

7.2.1. Expressing and verifying properties of multi-agent systems

Participants : Ocan Sankur, Nicolas Markey

7.2.1.1. Admissible strategies in controller synthesis.

In game theory, a strategy is dominated by another one if the latter systematically yields a payoff as good as the former, while also yielding a better payoff in some cases. A strategy is admissible if it is not dominated. This notion is well-studied in game theory and is useful to describe the set of strategies that are "reasonable" (i.e., whose choice can be justified; here, no players would play a dominated strategy, since better strategies exist). Recent works studied this notion in graph games with omega-regular objectives and investigated its applications in controller synthesis. For multi-agent controller synthesis, admissibility can be used as a hypothesis on the behaviors of each agent, thus enabling a compositional reasoning framework for controller synthesis.

We continue the study of admissibility in controller synthesis with three developments detailed as follows:

- In [29], we study the characterization and computation of admissible strategies in multiplayer concurrent games. We study both deterministic strategies and randomized ones with almost-sure winning criteria. We prove that admissible strategies always exist in concurrent games, and we characterise them precisely. Then, when the objectives of the players are omega-regular, we show how to perform assume-admissible synthesis, i.e., how to compute admissible strategies that win (almost surely) under the hypothesis that the other players play admissible strategies only.
- In [30], we study timed games, which are multiplayer games played on arena defined by timed automata, which are a particular case of concurrent games. First, we show that admissible strategies may not exist in timed games with a continuous semantics of time, even for safety objectives. Second, we show that the discrete time semantics of timed games is better behaved w.r.t. admissibility: the existence of admissible strategies is guaranteed in that semantics. Third, we provide symbolic algorithms to solve the model-checking problem under admissibility and the assume-admissible synthesis problem for real-time non-zero sum n-player games for safety objectives.
- In [26], we study admissible strategies in games with imperfect information. We show that in stark contrast with the perfect information variant, admissible strategies are only guaranteed to exist when players have objectives that are closed sets. As a consequence, we also study decision problems related to the existence of admissible strategies for regular games as well as finite duration games.

7.2.1.2. Strategy dependences in Strategy Logic.

Strategy Logic (SL) is a very expressive logic for specifying and verifying properties of multi-agent systems: in SL, one can quantify over strategies, assign them to agents, and express properties of the resulting plays (using linear-time temporal logic). This defines a very expressive framework, encompassing e.g. (pure) Nash equilibria, or admissibility. Such a powerful framework has two drawbacks: first, SL model checking has non-elementary complexity; second, the exact semantics of SL is rather intricate, and may not correspond to what is expected.

In [49], we focus on *strategy dependences* in SL, by tracking how existentially-quantified strategies in a formula may (or may not) depend on other strategies selected in the formula. We study different kinds of dependences, refining a previous approach [52], and prove that they give rise to different satisfaction relations. In the setting where strategies may only depend on what they have observed, we identify a large fragment of SL for which we prove model checking can be performed in *2EXPTIME*.

7.2.2. Active diagnosis

Participants : Nathalie Bertrand, Blaise Genest, Engel Lefaucheux

7.2.2.1. Diagnosis and control of the degradation of probabilistic systems.

Active diagnosis is performed by a controler so that a system becomes diagnosable. In order to avoid the controler to degrade the funtionning of the system too much, one often provides it with an additional objective specifying the desired quality of service.

In the context of probabilistic systems, a possible specification consists in requiring a positive probability of infinite correct runs, referred to as the safe active diagnosis. In [42], we introduced two alternative specifications. First (γ, v)-correction of a system associates with an execution a correction value which depends on a discount factor γ, and the controler must ensure an expected correction value greater than a threshold v. Second, α-persistence requires that asymptotically, at each time unit, a proportion at least α of runs that were correct so far remain correct.

Our contributions are twofold. On the one hand, from a semantical viewpoint, we make explicit the equivalences and (non-)implications between the various notions, for finite-state systems as well as infinite-state ones. On the other hand, algorithmically, we establish the decidability frontier of the corresponding decision problems, and for decidable problems characterize their precise complexity, together with algorithms to design controlers.

7.2.2.2. Probabilistic Disclosure: Maximisation vs. Minimisation.

We consider opacity questions where an observation function provides to an external attacker a view of the states along executions and secret executions are those visiting some secret state from a fixed subset. Disclosure occurs when the observer can deduce from a finite observation that the execution is secret. In a probabilistic and non deterministic setting, where an internal agent can choose between actions, there are two points of view, depending on the status of this agent: the successive choices can either help the attacker trying to disclose the secret, if the system has been corrupted, or they can prevent disclosure as much as possible if these choices are part of the system design. In the former situation, corresponding to a worst case, the disclosure value is the supremum over the strategies of the probability to disclose the secret (maximisation), whereas in the latter case, the disclosure is the infimum (minimisation). We address quantitative problems (relation between the optimal value and a threshold) and qualitative ones (when the threshold is zero or one) related to both forms of disclosure for a fixed or finite horizon. For all problems, we characterise their decidability status and their complexity. Surprisingly, while in maximisation problems optimal strategies may be chosen among deterministic ones, it is not the case for minimisation problems, but more minimisation problems than maximisation ones are decidable. These results appeared in [36].

7.2.3. Control and enforcement for quantitative systems

Participants : Nathalie Bertrand, Blaise Genest, Thierry Jéron, Hervé Marchand, Nicolas Markey

7.2.3.1. Qualitative determinacy and Decidability of Stochastic Games with Signals.

In [17], we consider two-person zero-sum stochastic games with signals, a standard model of stochastic games with imperfect information. The only source of information for the players consists of the signals they receive; they cannot directly observe the state of the game, nor the actions played by their opponent, nor their own actions.

We are interested in the existence of almost-surely winning or positively winning strategies, under reachability, safety, Büchi, or co-Büchi winning objectives, and the computation of these strategies when the game has finitely many states and actions. We prove two qualitative determinacy results. First, in a reachability game, either player 1 can achieve almost surely the reachability objective, or player 2 can achieve surely the dual safety objective, or both players have positively winning strategies. Second, in a Büchi game, if player 1 cannot achieve almost surely the Büchi objective, then player 2 can ensure positively the dual co-Büchi objective. We prove that players only need strategies with finite memory. The number of memory states needed to win with finite-memory strategies ranges from one (corresponding to memoryless strategies) to doubly exponential, with matching upper and lower bounds. Together with the qualitative determinacy results, we also provide fix-point algorithms for deciding which player has an almost-surely winning or a positively winning strategy and for computing an associated finite-memory strategy. Complexity ranges from *EXPTIME* to *2EXPTIME*, with matching lower bounds. Our fix-point algorithms also enjoy a better complexity in the cases where one of the players is better informed than their opponent.

Our results hold even when players do not necessarily observe their own actions. The adequate class of strategies, in this case, is mixed or general strategies (they are equivalent). Behavioral strategies are too restrictive to guarantee determinacy: it may happen that one of the players has a winning general strategy but none of them has a winning behavioral strategy. On the other hand, if a player can observe their actions, then general, mixed, and behavioral strategies are equivalent. Finite-memory strategies are sufficient for determinacy to hold, provided that randomized memory updates are allowed.

7.2.3.2. Average-energy games.

In [34], we consider average-energy games, where the goal is to minimize the long-run average of the accumulated weight (seen as an *energy level*) in a two-player game on a finite-state weighted automaton. Decidability of average-energy games with a lower-bound constraint on the energy level (but no upper bound) is an open problem; in particular, there is no known upper bound on the memory that is required for winning strategies.

By reducing average-energy games with lower-bounded energy to infinite-state mean-payoff games and analyzing the frequency of low-energy configurations, we show an almost tight doubly-exponential upper bound on the necessary memory, and that the winner of average-energy games with lower-bounded energy can be determined in doubly-exponential time. We also prove *EXPSPACE*-hardness of this problem.

Finally, we consider multi-dimensional extensions of all types of average-energy games: without bounds, with only a lower bound, and with both a lower and an upper bound on the energy. We show that the fully-bounded version is the only case to remain decidable in multiple dimensions.

7.2.3.3. Runtime enforcement.

The journal paper [23] details our work about predictive runtime enforcement, done in collaboration with University Aalto (Finland) and Inria CORSE/LIG Grenoble.

Runtime enforcement (RE) is a technique to ensure that the (untrustworthy) output of a black-box system satisfies some desired properties. In RE, the output of the running system, modeled as a sequence of events, is fed into an enforcer. The enforcer ensures that the sequence complies with a certain property, by delaying or modifying events if necessary. This paper deals with predictive runtime enforcement, where the system is not entirely black-box, but we know something about its behavior. This a priori knowledge about the system allows to output some events immediately, instead of delaying them until more events are observed, or even blocking them permanently. This in turn results in better enforcement policies. We also show that if we have no knowledge about the system, then the proposed enforcement mechanism reduces to standard (non-predictive) runtime enforcement. All our results related to predictive RE of untimed properties are also formalized and proved in the Isabelle theorem prover. We also discuss how our predictive runtime enforcement framework can be extended to enforce timed properties.

The journal paper [24], done in collaboration with LaBRI Bordeaux and Inria Corse/LIG Grenoble, deals with runtime enforcement of untimed and timed properties with uncontrollable events. Runtime enforcement consists in defining and using mechanisms that modify the executions of a running system to ensure their

correctness with respect to a desired property. We introduce a framework that takes as input any regular (timed) property described by a deterministic automaton over an alphabet of events, with some of these events being uncontrollable. An uncontrollable event cannot be delayed nor intercepted by an enforcement mechanism. Enforcement mechanisms should satisfy important properties, namely soundness, compliance, and optimality—meaning that enforcement mechanisms should output as soon as possible correct executions that are as close as possible to the input execution. We define the conditions for a property to be enforceable with uncontrollable events. Moreover, we synthesise sound, compliant, and optimal descriptions of runtime enforcement mechanisms at two levels of abstraction to facilitate their design and implementation.

7.2.3.4. Control of logico-numerical systems.

In paper [32], we have targeted the problem of the safe control of reconfigurations in component-based software systems, where strategies of adaptation to variations in both their environment and internal resource demands need to be enforced. In this context, the computing system involves software components that are subject to control decisions. We have approached this problem under the angle of discrete-event systems (DES), involving properties on events observed during the execution (e.g., requests of computing tasks, work overload), and a state space representing different configurations such as activity or assemblies of components. We have considered in particular the potential of applying novel logico-numerical control techniques to extend the expressivity of control models and objectives, thereby extending the application of DES in component-based software systems. We elaborate methodological guidelines for the application of logico-numerical control based on a case-study, and validate the result experimentally.

7.2.4. Smart regulation for urban trains

Participants : Éric Fabre, Loïc Hélouët, Hervé Marchand, Karim Kecir

The regulation of subway lines consists in accomodating small random perturbations in transit times as well as more impacting incidents, by playing on continuous commands (transit times and dwell times) and by making more complex decisions (insertions or extractions of trains, changes of missions, overpassing, shorter returns, etc.) The objectives are multiple: ensuring the regularity and punctuality of trains, adapting to transportation demand, minimizing energy consumption, etc. We have developed an event-based control strategy that aims at equalizing headways on a line. This distributed control strategy is remarquably robust to perturbations and reactive enough to accomodate train insertions/extractions. We have integrated this control startegy to our SIMSTORS software. We have also developed another approach based on event graphs in order to optimally interleave trains at a junction. We started investigating new predictive control policies based of optimisation of criteria in forecast schedules [43].

In [47], we have extended a work started in 2016, that considers realizability of schedules by metro systems. Schedules are defined as high-level views of desired executions of systems, and represented as partial orders decorated with timing constraints. Train networks are modeled as stochastic time Petri nets (STPN) with an elementary (1-bounded) semantics. We have proposed a notion of time processes to give a partial-order semantics to STPNs. We then have considered Boolean realizability: a schedule S is realizable by a net N if S embeds in a time process of N that satisfies all its constraints. However, with continuous time domains, the probability of a time process with exact dates is null. We thus consider probabilistic realizability up to α time units, that holds if the probability that N realizes S with constraints enlarged by α is strictly positive. Upon a sensible restriction guaranteeing time progress, Boolean and probabilistic realizability of a schedule can be checked on the finite set of symbolic prefixes extracted from a bounded unfolding of the net. We give a construction technique for these prefixes and show that they represent all time processes of a net occurring up to a given maximal date. We then show how to verify existence of an embedding and compute the probability of its realization. The technique has then been illustrated by a concrete example, namely deciding wheter a simple flip-flop shunting mechanism suffices to route trains in appropriate direction when delays can occur in trips or during stops at stations. We have also conducted a series of experiment [28] with the SIMSTORS tool to obtain statistics, and show feasibiliy of Key Performance Indicators (KPIs) evaluation with this formal model.

A second line of research relates to the development of new regulation strategies. New techniques were derived to equalize headways of trains along a line, and thus improve regularity and resilience to perturbations. A distributed control strategy was developed, easily implementable in existing rule engines. Simulations have proved the efficiency of this technique on orbital lines. We have also developed a global regulation approach based on timed event graphs. In this setting, control is event-based: a command is issued each time a train crosses a control point, but it takes into account information along the whole line and for a finite time horizon. This amounts to adapting the whole time-table for any new event in the system. This approach has been proved to perform well at junctions (on computer simulations), where randomly spaced trains arriving from two branches must be correctly interleaved at the junction of the two lines, while at the same time train intervals must be equalized in all branches. We are now working on the combinatorial aspects of the question, in order to reduce energy consumption (by synchronizing arrivals and departures of trains), and in order to allow for insertions/extractions and reorderings of trains.

Several patents are in preparation for this activity.

7.3. Management of Large Distributed Systems

7.3.1. Analysis and synthesis of distributed systems

Participants : Éric Badouel, Thierry Jéron, Hervé Marchand, The Anh Pham

7.3.1.1. Control of Distributed Systems.

In [40], we have extended our examination of decentralized discrete-event-system architectures that use exclusive or (XOR) as the fusion rule to reach control decisions. A characterization of XOR inference-observable languages has been provided. Additionally, XOR observability is defined for languages that are not inference-observable but are distributed-observable.

7.3.1.2. Verification of distributed applications

In the context of IPL HAC-SPECIS, in collaboation with Martin Quinson (Myriads Inria project team) we are interested in the verification of real distributed applications.

In the conference paper [38] we explain the current status of the tool SimGridMC used for the verification of MPI applications. SimGridMC (also dubbed Mc SimGrid) is a stateful Model Checker for MPI applications. It is integrated to SimGrid, a framework mostly dedicated to predicting the performance of distributed applications. We describe the architecture of McSimGrid, and show how it copes with the state space explosion problem using Dynamic Partial Order Reduction and State Equality algorithms. As case studies we show how SimGrid can enforce safety and liveness properties for MPI applications, as well as global invariants over communication patterns.

7.3.2. Analysis of parameterized systems

Participants : Nathalie Bertrand, Éric Fabre, Blaise Genest, Matthieu Pichené, Ocan Sankur

7.3.2.1. Parameterized Verification of a time-synchronization protocol.

In [41], we consider distributed timed systems that implement leader-election protocols, which are at the heart of clock-synchronization protocols. We develop abstraction techniques for parameterized model checking of such protocols under arbitrary network topologies, where nodes have independently-evolving clocks. We apply our technique for model checking the root election part of the flooding time-synchronisation protocol (FTSP), and obtain improved results compared to previous work. We model-check the protocol for all topologies in which the distance to the node to be elected leader is bounded by a given parameter.

7.3.2.2. Controlling population models.

In [33], we introduce a new setting where a population of agents, each modelled by a finite-state system, are controlled uniformly: the controller applies the same action to every agent. The framework is largely inspired by the control of a biological system, namely a population of yeasts, where the controller may only change the environment common to all cells. We study a synchronisation problem for such populations: no matter how individual agents react to the actions of the controller, the controller aims at driving all agents synchronously to a target state. The agents are naturally represented by a non-deterministic finite state automaton (NFA), the same for every agent, and the whole system is encoded as a 2-player game. The first player (Controller) chooses actions, and the second player (Agents) resolves non-determinism for each agent. The game with m agents is called the m-population game. This gives rise to a parameterized control problem (where control refers to 2-player games), namely the population control problem: can Controller control the m-population game for all $m \in \mathbb{N}$, whatever Agents does?

In this work, we prove that the population control problem is decidable, and it is an *EXPTIME*-complete problem. As far as we know, this is one of the first results on parameterized control. Our algorithm, not based on cut-off techniques, produces winning strategies which are symbolic, that is, they do not need to count precisely how the population is spread between states. We also show that if there is no winning strategy, then there is a population size M such that Controller wins the m-population game if, and only if, $m \leq M$. Surprisingly, M can be doubly-exponential in the number of states of the NFA, with tight upper and lower bounds.

7.3.2.3. Handling large biological systems.

This year, we propose to use approximated probabilistic distribution to handle large homogeneous populations of cells [39]. Beyond classical approximations, we propose to use the Chow-Liu tree representation, based on *non-disjoint* clusters of two variables. Our experiments show that our proposed approximation scheme is more accurate than existing ones to model probability distributions deriving from biopathways, while requiring a minimal complexity overhead.

To handle *dynamics* of a population of cells governed by biopathways, we develop *coarse-grained* abstractions of the biological pathways [21], and more precisely *Dynamic Bayesian Networks* (DBNs). We show that simulating a DBN is much faster than simulating the fine-grained model it abstracts, for comparable prediction performances.

We also explore the approximate inference problem of DBNs, that is, *computing* the probability distributions at every time point given the initial distribution at time 0. We evaluate several classical approximate inference algorithms for DBNs, and compare with a new method we propose, which consists in using the Chow-Liu tree approximation to represent distributions at each time step. It is very accurate, yet efficient according to experiments we report. We finally provide an error analysis of this approximate inference algorithm [39].

7.4. Data-Driven Systems

7.4.1. Incremental process discovery using Petri-net synthesis.

Participants : Éric Badouel

In [16], we present an incremental process discovery using Petri-net synthesis. Process discovery aims at constructing a model from a set of observations given by execution traces (a log). Petri nets are a preferred target model in that they produce a compact description of the system by exhibiting its concurrency. This article presents a process-discovery algorithm using Petri-net synthesis, based on the notion of region introduced by A. Ehrenfeucht and G. Rozenberg, and using techniques from linear algebra. The algorithm proceeds in three successive phases which make it possible to find a compromise between the ability to infer behaviours of the system from the set of observations while ensuring a parsimonious model, in terms of fitness, precision and simplicity. All used algorithms are incremental which means that one can modify the produced model when new observations are reported without reconstructing the model from scratch.

7.4.2. An artifact model with imprecision and uncertainty

Participants : Éric Badouel, Loïc Hélouët

In the context of the HeadWork ANR project, we started investigating how complex workflows can be defined to handle uncertainty, and use joint knowledge of pools of user to build correct information. The solution proposed so far is a variant of business artifact managing fuzzy datasets. As there are several ways to reach an acceptable final and sufficiently precise dataset, we started investigating equivalence of complex workflows with partial information to allow refinement, enhance performance of data collection, with mastered precision loss.

TOCCATA Project-Team

7. New Results

7.1. Deductive Verification

Synthetic topology in HoTT for probabilistic programming. F. Faissole and B. Spitters have developed a mathematical formalism based on synthetic topology and homotopy type theory to interpret probabilistic algorithms. They suggest to use proof assistants to prove such programs [39] [31]. They also have formalized synthetic topology in the Coq proof assistant using the HoTT library. It consists of a theory of lower reals, valuations and lower integrals. All the results are constructive. They apply their results to interpret probabilistic programs using a monadic approach [28].

Defunctionalization for proving higher-order programs. J.-C. Filliâtre and M. Pereira proposed a new approach to the verification of higher-order programs, using the technique of defunctionalization, that is, the translation of first-class functions into first-order values. This is an early experimental work, conducted on examples only within the Why3 system. This work was published at JFLA 2017 [29].

Extracting Why3 programs to C programs. R. Rieu-Helft, C. Marché, and G. Melquiond devised a simple memory model for representing C-like pointers in the Why3 system. This makes it possible to translate a small fragment of Why3 verified programs into idiomatic C code [30]. This extraction mechanism was used to turn a verified Why3 library of arbitrary-precision integer arithmetic into a C library that can be substituted to part of the GNU Multi-Precision (GMP) library [23].

Verification of highly imperative OCaml programs with Why3 J.-C. Filliâtre, M. Pereira and S. Melo de Sousa proposed a new methodology for proving highly imperative OCaml programs with Why3. For a given OCaml program, a specific memory model is built and one checks a Why3 program that operates on it. Once the proof is complete, they use Why3's extraction mechanism to translate its programs to OCaml, while replacing the operations on the memory model with the corresponding operations on mutable types of OCaml. This method is evaluated on several examples that manipulate linked lists and mutable graphs [20].

7.2. Automated Reasoning

A Three-tier Strategy for Reasoning about Floating-Point Numbers in SMT. The SMT-LIB standard defines a formal semantics for a theory of floating-point (FP) arithmetic (FPA). This formalization reduces FP operations to reals by means of a rounding operator, as done in the IEEE-754 standard. Closely following this description, S. Conchon, M. Iguernlala, K. Ji, G. Melquiond and C. Fumex propose a three-tier strategy to reason about FPA in SMT solvers. The first layer is a purely axiomatic implementation of the automatable semantics of the SMT-LIB standard. It reasons with exceptional cases (e.g. overflows, division by zero, undefined operations) and reduces finite representable FP expressions to reals using the rounding operator. At the core of the strategy, a second layer handles a set of lemmas about the properties of rounding. For these lemmas to be used effectively, the instantiation mechanism of SMT solvers is extended to tightly cooperate with the third layer, the NRA engine of SMT solvers, which provides interval information. The strategy is implemented in the Alt-Ergo SMT solver and validated on a set of benchmarks coming from the SMT-LIB competition, and also from the deductive verification of C and Ada programs. The results show that the approach is promising and compete with existing techniques implemented in state-of-the-art SMT solvers. This work was presented at the CAV conference [18].

Lightweight Approach for Declarative Proofs. M. Clochard designed an extension of first-order logic, for describing reasoning steps needed to discharge a proof obligation. The extension is under the form of two new connectives, called proof indications, that allow the user to encode reasoning steps inside a logic formula. This extension makes possible to use the syntax of formulas as a proof language. The approach was presented at the JFLA conference [26] and implemented in Why3. It brings a lightweight mechanism for declarative proofs in an environment like Why3 where provers are used as black boxes. Moreover, this mechanism restricts the scope of auxiliary lemmas, reducing the size of proof obligations sent to external provers.

7.3. Certification of Algorithms, Languages, Tools and Systems

Formalization and closedness of finite dimensional subspaces. F. Faissole formalized a theory of finite dimensional subspaces of Hilbert spaces in order to apply the Lax-Milgram Theorem on such subspaces. He had to prove, in the Coq proof assistant, that finite dimensional subspaces of Hilbert spaces are closed in the context of general topology using filters [19]. He also formalized both finite dimensional modules and finite dimensional subspaces of modules. He compared the two formalizations and showed a complementarity between them. He proved that the product of two finite dimensional modules is a finite dimensional module [27].

Verified numerical approximations of improper definite integrals. The CoqInterval library provides some tactics for computing and formally verifying numerical approximations of real-valued expressions inside the Coq system. In particular, it is able to compute reliable bounds on proper definite integrals [111]. A. Mahboubi, G. Melquiond, and T. Sibut-Pinote extended these algorithms to also cover some improper integrals, e.g., those with an unbounded integration domain [40]. This makes CoqInterval one of the very few tools able to produce reliable results for improper integrals, be they formally verified or not.

A Coq Formal Proof of the Lax–Milgram theorem. S. Boldo, F. Clément, F. Faissole, V. Martin, and M. Mayero worked on a Coq formal proof of the Lax–Milgram theorem. It is one of the theoretical cornerstone for the correctness of the Finite Element Method. It required many results from linear algebra, geometry, functional analysis, and Hilbert spaces [13] [24].

Formalization of numerical filters S. Boldo, D. Gallois-Wong, and T. Hilaire developped a formalization in the Coq proof assistant of numerical filters. It includes equivalences between several expressions and the formal proof of the Worst-Case Peak Gain Theorem to bound the magnitude of the outputs (and every intern variable) of stable filters.

A Verified OCaml Library. Abstract Libraries are the basic building blocks of any realistic programming project. It is thus of utmost interest for a programmer to build her software on top of bug-free libraries. At the ML family workshop [38], A. Charguéraud, J.-C. Filliâtre, M. Pereira and F. Pottier presented the ongoing VOCAL project, which aims at building a mechanically verified library of general-purpose data structures and algorithms, written in the OCaml language. A key ingredient of VOCAL is the design of a specification language for OCaml, independently of any verification tool.

Formal Analysis of shell scripts. The shell language is widely used for various system administration tasks on UNIX machines. The CoLiS project aims at applying formal methods for verifying scripts used for installation of packages of software distributions. The syntax and semantics of shell are particularly treacherous. They proposed a new language called CoLiS which, on the one hand, has well-defined static semantics and avoids some of the pitfalls of the shell, and, on the other hand, is close enough to the shell to be the target of an automated translation of the scripts in our corpus. In collaboration with N. Jeannerod and R. Treinen, C. Marché formalized the syntax and semantics of CoLiS in Why3, defined an interpreter for the language in the WhyML programming language, and present an automated proof in the Why3 proof environment of soundness and completeness of this interpreter with respect to the formal semantics [22]. The development is available in Toccata's gallery http://toccata.lri.fr/gallery/colis_interpreter.en.html. This formalized interpreter is extracted to OCaml and the verified code is integrated into a prototype software toolset developed by I. Dami and C. Marché [36].

A verified yet efficient arbitrary-precision integer library. R. Rieu-Helft used the Why3 system to implement, specify, and verify a library of arbitrary-precision integer arithmetic: comparison, addition, multiplication, shifts, division. A lot of efforts were put into replicating and verifying the numerous implementation tricks the GMP library uses to achieve state-of-the-art performances, especially for the division algorithm. While the resulting library is nowhere near as fast as the hand-written assembly code GMP uses, it is competitive with the generic C code of GMP for small integers (i.e., mini-GMP) [23]. The development is available in Toccata's gallery http://toccata.lri.fr/gallery/multiprecision.en.html.

Case study: algorithms for matrix multiplication. M. Clochard, L. Gondelman and M. Pereira worked on a case study about matrix multiplication. Two variants for the multiplication of matrices are proved: a naive version using three nested loops and Strassen's algorithm. To formally specify the two multiplication algorithms, they developed a new Why3 theory of matrices, and they applied a reflection methodology to conduct some of the proofs. A first version of this work was presented at the VSTTE Conference in 2016 [78]. An extended version that considers arbitrary rectangular matrices instead of square ones is published in the Journal of Automated Reasoning [12]. The development is available in Toccata's gallery http://toccata.lri.fr/gallery/verifythis_2016_matrix_multiplication.en.html.

Case studies: Strongly Connected Components in Directed Graphs As part of a larger set of case studies on algorithms on graphs http://pauillac.inria.fr/~levy/why3/, R. Chen and J.-J. Lévy work on formal verification of algorithms for computing strongly connected components of directed graphs. The formal proofs are conducted using Why3. The formal proof of Tarjan's algorithm was presented at the French-speaking symposium JFLA 2017 [25] and then at the VSTTE 2017 international conference [17]

A Formally Proved, Complete Algorithm for Path Resolution with Symbolic Links In the context of file systems like those of Unix, path resolution is the operation that given a character string denoting an access path, determines the target object (a file, a directory, etc.) designated by this path. This operation is not trivial because of the presence of symbolic links. Indeed, the presence of such links may induce infinite loops in the resolution process. R. Chen, M. Clochard and C. Marché consider a path resolution algorithm that always terminates, detecting if it enters an infinite loop and reports a resolution failure in such a case. They propose a formal specification of path resolution and they formally prove that their algorithm terminates on any input, and is correct and complete with respect to this formal specification. [11]. The development is available in Toccata's gallery http://toccata.lri.fr/gallery/path_resolution.en.html.

7.4. Floating-Point and Numerical Programs

Computer Arithmetic and Formal Proofs: Verifying Floating-point Algorithms with the Coq System S. Boldo and G. Melquiond published a book that provides a comprehensive view of how to formally specify and verify tricky floating-point algorithms with the Coq proof assistant. It describes the Flocq formalization of floating-point arithmetic and some methods to automate theorem proofs. It then presents the specification and verification of various algorithms, from error-free transformations to a numerical scheme for a partial differential equation. The examples cover not only mathematical algorithms but also C programs as well as issues related to compilation [32].

Automating the Verification of Floating-Point Programs. The level of proof success and proof automation highly depends on the way the floating-point operations are interpreted in the logic supported by back-end provers. C. Fumex, C. Marché and Y. Moy addressed this challenge by combining multiple techniques to separately prove different parts of the desired properties. They use abstract interpretation to compute numerical bounds of expressions, and use multiple automated provers, relying on different strategies for representing floating-point computations. One of these strategies is based on the native support for floating-point arithmetic recently added in the SMT-LIB standard. The approach is implemented in the Why3 environment and its front-end SPARK 2014.

It is validated experimentally on several examples originating from industrial use of SPARK 2014 [37], [21].

Round-off Error Analysis of Explicit One-Step Numerical Integration Methods. S. Boldo, A. Chapoutot, and F. Faissole provided bounds on the round-off errors of explicit one-step numerical integration methods, such as Runge-Kutta methods. They developed a fine-grained analysis that takes advantage of the linear stability of the scheme, a mathematical property that vouches the scheme is well-behaved [14].

Robustness of 2Sum and Fast2Sum. S. Boldo, S.Graillat, and J.-M. Muller worked on the 2Sum and Fast2Sum algorithms, that are important building blocks in numerical computing. They are used (implicitly or explicitly) in many compensated algorithms or for manipulating floating-point expansions. They showed that these algorithms are much more robust than it is usually believed: the returned result makes sense even when the rounding function is not round-to-nearest, and they are almost immune to overflow [10].

Formal Verification of a Floating-Point Expansion Renormalization Algorithm. Many numerical problems require a higher computing precision than the one offered by standard floating-point formats. A common way of extending the precision is to use floating-point expansions. S. Boldo, M. Joldes, J.-M. Muller, and V. Popescu proved one of the algorithms used as a basic brick when computing with floating-point expansions: renormalization that “compresses” an expansion [15].

VERIDIS Project-Team

7. New Results

7.1. Automated and Interactive Theorem Proving

Participants: Haniel Barbosa, Jasmin Christian Blanchette, Martin Bromberger, Simon Cruanes, Daniel El Ouraoui, Mathias Fleury, Pascal Fontaine, Stephan Merz, Martin Riener, Hans-Jörg Schurr, Martin Strecker, Thomas Sturm, Andreas Teucke, Sophie Tourret, Marco Voigt, Tung Vu Xuan, Uwe Waldmann, Daniel Wand, Christoph Weidenbach.

7.1.1. IsaFoL: Isabelle Formalization of Logic

Joint work with Andreas Halkjær From (DTU Copenhagen), Alexander Birch Jensen (DTU Copenhagen), Maximilian Kirchmeier (TU München), Peter Lammich (TU München), John Bruntse Larsen (DTU Copenhagen), Julius Michaelis (TU München), Tobias Nipkow (TU München), Nicolas Peltier (IMAG Grenoble) Anders Schlichtkrull (DTU Copenhagen), Dmitriy Traytel (ETH Zürich), and Jørgen Villadsen (DTU Copenhagen).

Researchers in automated reasoning spend a significant portion of their work time specifying logical calculi and proving metatheorems about them. These proofs are typically carried out with pen and paper, which is error-prone and can be tedious. As proof assistants are becoming easier to use, it makes sense to employ them.

In this spirit, we started an effort, called IsaFoL (Isabelle Formalization of Logic), that aims at developing libraries and methodology for formalizing modern research in the field, using the Isabelle/HOL proof assistant.[0] Our initial emphasis is on established results about propositional and first-order logic. In particular, we are formalizing large parts of Weidenbach's forthcoming textbook, tentatively called *Automated Reasoning—The Art of Generic Problem Solving*.

The objective of formalization work is not to eliminate paper proofs, but to complement them with rich formal companions. Formalizations help catch mistakes, whether superficial or deep, in specifications and theorems; they make it easy to experiment with changes or variants of concepts; and they help clarify concepts left vague on paper.

The repository contains 14 completed entries and four entries that are still in development. Notably, Mathias Fleury formalized a SAT solver framework with learn, forget, restart, and incrementality. This year he extended it with key optimizations such as the two-watched-literal procedure. The corresponding paper, written together with Jasmin Blanchette and Peter Lammich, was accepted at a highly competitive conference (CPP 2018).

7.1.2. Extension of Term Orders to λ-Free Higher-Order Logic

Superposition is one of the most successful proof calculi for first-order logic today, but in contrast to resolution, tableaux, and connections, it has not yet been generalized to higher-order logic (also called simple type theory). Yet, most proof assistants and many specification languages are based on some variant of higher-order logic.

This motivates us to design a *graceful* generalization of superposition: a proof calculus that behaves like standard superposition on first-order problems and that smoothly scales up to arbitrary higher-order problems. A challenge is that superposition relies on a simplification order, which is fixed in advance of the proof attempt, to prune the search space.

[0] https://bitbucket.org/isafol/isafol/wiki/Home

We started our investigations by focusing on a fragment devoid of λ-abstractions, but with partial application and application of variables, two crucial higher-order features. We generalized the two main orders that are used in superposition-based provers today—the lexicographic path order (LPO) [27] and the Knuth-Bendix order (KBO) [21]. The new orders gracefully generalize their first-order counterparts and enjoy nearly all properties needed for superpositions. An exception is compatibility with contexts, which is missing for LPO and some KBO variants. Preliminary work suggests that we can define a version of the superposition calculus that works well in theory and practice (i.e., is refutationally complete and does not lead to a search-space explosion) despite the missing property.

7.1.3. A Fine-Grained Approach of Understanding First-Order Logic Complexity

By the introduction of the separated fragment [65] we have initiated a new framework for a fine-grained understanding of the complexity of fragments of first-order logic, with and without the addition of theories. We have related the classes of the polynomial hierarchy to subclasses of the separated fragment [40] and developed new decidability results [36], [41] based on the techniques of our framework for the combination of the Bernays-Schoenfinkel subfragment with linear arithmetic.

7.1.4. Theorem Proving Based on Approximation-Refinement into the Monadic Shallow Linear Fragment with Straight Dismatching Constraints

We have introduced an approximation-refinement approach for first-order theorem proving based on counterexample-guided abstraction refinement [39]. A given first-order clause set is transformed into an over-approximation contained in the fragment of monadic, shallow, linear clauses with straight dismatching constraints. We have shown the fragment to be decidable, strictly extending known results. If the abstraction obtained that way is satisfiable, so is the original clause set. However, if it is unsatisfiable, then the approximation provides a terminology for lifting the found refutation, step by step, into a proof for the original clause set. If lifting fails, the cause is analyzed to refine the original clause set such that the found refutation is ruled out for the future, and the procedure repeats. We have shown that this approach is superior to all known calculi on certain classes of first-order clauses. In particular, it is able to detect satisfiability of clause sets that have only infinite models.

7.1.5. Combination of Satisfiability Procedures

Joint work with Christophe Ringeissen from the PESTO project-team of Inria Nancy – Grand Est, and Paula Chocron at IIIA-CSIC, Bellaterra, Catalonia, Spain.

A satisfiability problem is often expressed in a combination of theories, and a natural approach consists in solving the problem by combining the satisfiability procedures available for the component theories. This is the purpose of the combination method introduced by Nelson and Oppen. However, in its initial presentation, the Nelson-Oppen combination method requires the theories to be signature-disjoint and stably infinite (to ensure the existence of an infinite model). The design of a generic combination method for non-disjoint unions of theories is difficult, but it is worth exploring simple non-disjoint combinations that appear frequently in verification. An example is the case of shared sets, where sets are represented by unary predicates. Another example is the case of bridging functions between data structures and a target theory (e.g., a fragment of arithmetic).

In 2015, we defined [55] a sound and complete combination procedure à la Nelson-Oppen for the theory of absolutely free data structures (including lists and trees) connected to another theory via bridging functions. This combination procedure has also been refined for standard interpretations. The resulting theory has a nice politeness property, enabling combinations with arbitrary decidable theories of elements. We also investigated other theories [56] amenable to similar combinations: this class includes the theory of equality, the theory of absolutely free data structures, and all the theories in between.

In 2017, we have been improving the framework and unified both results. A new paper is in preparation.

7.1.6. Quantifier Handling in SMT

Joint work with Andrew J. Reynolds, Univ. of Iowa, USA.

SMT solvers generally rely on various instantiation techniques for handling quantifiers. We built a unifying framework encompassing quantified formulas with equality and uninterpreted functions, such that the major instantiation techniques in SMT solving can be cast in that framework. It is based on the problem of E-ground (dis)unification, a variation of the classic Rigid E-unification problem. We introduced a sound and complete calculus to solve this problem in practice: Congruence Closure with Free Variables (CCFV). Experimental evaluations of implementations of CCFV demonstrate notable improvements in the state-of-the-art solver CVC4 and make the solver veriT competitive with state-of-the-art solvers for several benchmark libraries, in particular those originating in verification problems. This was the subject of a publication [20]. In later, unpublished work, we are revisiting enumerative instantiation for SMT. This effort takes place in the context of the Matryoshka project.

7.1.7. Non-Linear Arithmetic in SMT

In the context of the SMArT ANR-DFG (Satisfiability Modulo Arithmetic Theories), KANASA and SC^2 projects (cf. sections 9.1 and 9.3), we study the theory, design techniques, and implement software to push forward the non-linear arithmetic (NLA) reasoning capabilities in SMT. This year, we designed a framework to combine interval constraint propagation with other decision procedures for NLA, with promising results, notably in the international competition of SMT solvers SMT-COMP 2017. We also studied integration of these procedures into combinations of theories. The ideas are validated within the veriT solver, together with code from the raSAT solver (from JAIST). An article is in preparation.

We also adapted the subtropical method to use in an SMT context, with valuable results. This was the subject of a publication in 2017 [33].

7.1.8. Proofs for SMT

We have developed a framework for processing formulas in automatic theorem provers, with generation of detailed proofs. The main components are a generic contextual recursion algorithm and an extensible set of inference rules. Clausification, skolemization, theory-specific simplifications, and expansion of 'let' expressions are instances of this framework. With suitable data structures, proof generation adds only a linear-time overhead, and proofs can be checked in linear time. We implemented the approach in the SMT solver veriT. This allowed us to dramatically simplify the code base while increasing the number of problems for which detailed proofs can be produced, which is important for independent checking and reconstruction in proof assistants. This was the subject of a publication in [19]. This effort takes place in the context of the Matryoshka project.

7.1.9. Coding Modal and Description Logics in SAT solvers

The application scenario behind this research is the verification of graph transformations, which themselves are relevant for a wide range of practical problems such as pointer structures in imperative programs, graph databases or access control mechanisms.

Graph structures can typically be perceived as models of modal logics, and modal logics and variants (such as description logics that are the basis for the web ontology language OWL) are in principle suitable specification formalisms for graph transformations. It turns out, however, that pure modal logics are often not sufficiently expressive for the intended verification purpose and that extensions are needed for which traditional proof methods such as tableau calculi become complex: the termination of the calculi are often very difficult to prove, and huge efforts are required to obtain an efficient implementation.

For these reasons, we have explored methods of encoding the above-mentioned logics in SAT and SMT solvers such as CVC4 and veriT. The idea is to traverse the formula to be verified in order to span up a pre-model that possibly contains more elements (worlds in a Kripke structure) than the real model, and then to run a solver to find out which of these elements can effectively be realized. A prototype has been implemented, with encouraging results. It remains to connect this prototype to the graph verification engine and to publish this work.

7.1.10. Work on the TLA+ Proof System

We continued our work on encoding set-theoretic formulas in multi-sorted first-order logic, and in particular for SMT solvers. Specifically, we unified and streamlined a technique combining an injection of unsorted expressions into sorted languages, simplification by rewriting, and abstraction that underlies the SMT backend of the TLA$^+$ proof system TLAPS. A presentation of our technique was accepted in the journal *Science of Computer Programming*, to appear in 2018.

The proof of the join protocol in a pure-join variant of the Pastry protocol [63] implementing a distributed hash table over a peer-to-peer network is the largest case study carried out so far within TLAPS. Consisting of roughly 30k lines of proof, it was developed as part of Noran Azmy's PhD thesis, defended at the end of 2016 [51]. A presentation of the design of the protocol and its proof was accepted in the journal *Science of Computer Programming*, to appear in 2018.

7.1.11. Automated Analysis of Systems of ODE for Multistationarity

Joint work with R. Bradford and J. Davenport (Bath, UK), M. England (Coventry, UK), H. Errami, C. Hoyt, and A. Weber (Bonn, Germany), V. Gerdt (Dubna, Russia), D. Grigoriev (Lille, France), O. Radulescu (Montpellier, France)

We considered the problem of determining multiple steady states for positive real values in models of biological networks. Investigating the potential for these in models of the mitogen-activated protein kinases (MAPK) network has consumed considerable effort using special insights into the structure of corresponding models. We have applied combinations of symbolic computation methods for mixed equality/inequality systems, specifically automated deduction methods like virtual substitution, lazy real triangularization and cylindrical algebraic decomposition. We have determined multistationarity of an 11-dimensional MAPK network when numeric values are known for all but potentially one parameter. More precisely, our considered model has 11 equations in 11 variables and 19 parameters, 3 of which are of interest for symbolic treatment, and furthermore positivity conditions on all variables and parameters [28].

Subsequent work [31] demonstrates that our techniques benefit tremendously from a new graph theoretical symbolic preprocessing method. We apply our combined techniques to visualize of parameter regions for multistationarity. Comparing computation times and quality of results it turns out that our automated deduction-based approach clearly outperforms established numerical continuation methods.

While automated deduction technology is a bit under the hood here, this interdisciplinary research line addresses important questions related to contemporary research in systems biology. With researchers from that area very actively involved, the results are recognized also within their communities.

7.2. Formal Methods for Developing and Analyzing Algorithms and Systems

Participants: Marie Duflot-Kremer, Margaux Duroeulx, Souad Kherroubi, Poonam Kumari, Dominique Méry, Stephan Merz, Nicolas Schnepf, Christoph Weidenbach.

7.2.1. Making Explicit Domain Knowledge in Formal System Development

Joint work with partners of the IMPEX project.

As explained in the description of the IMPEX project in section 9.1 , we advocate that formal modeling languages should explicitly represent the knowledge resulting from an analysis of the application domain, and that ontologies are good candidates for handling explicit domain knowledge. Our objective in doing so is to offer rigorous mechanisms for handling domain knowledge in design models.

We developed the notion of dependency for state-based models. Context-awareness is an important feature in system design. We argue that in proof systems and conceptual modelling this notion should be highlighted precisely. Since we focus on conceptual modelling, understandability and clarity are of high importance. We introduce a new definition [37] for proof context in state-based formalisms with an application to the Event-B modeling language. Furthermore, we introduce a dependency relation between two Event-B models. The contextualization of Event-B models is based on knowledge provided from domains that we classified into

constraints, hypotheses and dependencies. The dependency mechanism between two models makes it possible to structure the development of systems models, by organizing phases identified in the analyzed process. These ideas are inspired by work based on the modelling of situations in situation theory that emphasize capabilities of type theory with regard to situation modelling to represent knowledge. Our approach is illustrated on small case studies, and was validated on a development of design patterns for voting protocols.

7.2.2. Incremental Development of Systems and Algorithms

Joint work with Manamiary Bruno Andriamiarina, Neeraj Kumar Singh (IRIT, Toulouse), Rosemary Monahan (NUI Maynooth, Ireland), Zheng Cheng (LINA, Nantes), and Mohammed Mosbah (LaBRI, Bordeaux).

The development of distributed algorithms and, more generally, of distributed systems, is a complex, delicate, and challenging process. The approach based on refinement applies a design methodology that starts from the most abstract model and leads, in an incremental way, to a distributed solution. The use of a proof assistant gives a formal guarantee on the conformance of each refinement with the model preceding it.

Our main result during 2017 is the development of a proved-based pattern for integrating the local computation models and the Visidia platform [32].

7.2.3. Modeling Network Flows in View of Building Security Chains

Joint work with Rémi Badonnel and Abdelkader Lahmadi of the Madynes team of Inria Nancy – Grand Est.

We are working on the application of formal modeling and verification techniques in the area of network communications, and in particular for constructing security functions in a setting of software-defined networks (SDN). Concretely, Nicolas Schnepf defined an extension of the Pyretic language [58] taking into account both the control and the data planes of SDN controllers and implemented a translation of that extension to the input languages of the nuXmv model checker and of SMT solvers. This work was published at NetSoft 2017 [38].

Extending this approach, we have worked on inferring probabilistic finite-state automata models that represent network flows generated by Android applications. The objective is to exploit this representation for generating security chains that detect significant deviations from the behavior represented by the automata and can initiate protective actions. Comparing our models with automata produced by the state-of-the-art tools Invarimint and Synoptic, we obtain representations that are as succinct as those inferred by Invarimint, and significantly smaller than Synoptic, but that include information about transition probability, which Invarimint does not. This work was accepted for publication at NOMS 2018.

7.2.4. Satisfiability Techniques for Reliability Assessment

Joint work with Nicolae Brînzei at Centre de Recherche en Automatique de Nancy.

The reliability of complex systems is typically assessed using probabilistic methods, based on the probabilities of failures of individual components, relying on graphical representations such as fault trees or reliability block diagrams. Mathematically, the dependency of the overall system on the working status of its components is described by its Boolean-valued *structure function*, and binary decision diagrams (BDDs) have been used to construct a succinct representation of that function. We explore the use of modern satisfiability techniques as an alternative to BDD-based algorithms. In [30], we develop three different algorithms for computing minimal tie sets (i.e., component configurations that ensure that the system is functioning). Our algorithms are based on either conjunctive or disjunctive normal form representations of the structure function or on the Hasse diagram representing the configurations. These algorithms have been prototypically implemented in Python, and we are evaluating them on existing benchmarks in order to understand which algorithm works best for typical fault dependencies.

7.2.5. Statistical evaluation of the robustness of production schedules

Joint work with Alexis Aubry, Sara Himmiche, Pascale Marangé, and Jean-François Pétin at Centre de Recherche en Automatique de Nancy.

Finding a good schedule for a production system, especially when it is flexible and when several machines can perform the same operation on products, is a challenging and interesting problem. For a long time, operations research has provided state-of-the-art methods for optimizing scheduling problems. However, approaches based on Discrete Event Systems present interesting alternatives, especially when dealing with uncertainties on the demand or the production time. In this particular case, the flexibility of the automata-based modeling approach is really useful. Using probabilistic timed automata, we demonstrated [35] that statistical model checking can be used successfully for evaluating the robustness of a given schedule w.r.t. probabilistic variations of the processing time. We were thus able to compare different schedules based on their level of service (i.e., the probability that the system will complete the production process within a deadline slightly higher that the schedule time) and their sensitivity (the minimal deadline for which the level of service is greater than a given threshold) [42].

An interdisciplinary workshop on this topic was organized jointly with our colleagues of Centre de Recherche en Automatique and funded by Fédération Charles Hermite.

7.2.6. Using Cubicle for Verifying TLA+ Specifications

Cubicle [0] is a model checker for the verification of parameterized transition systems whose state is described by arrays of variables indexed by an abstract sort representing processes. During her internship, Poonam Kumari designed a translation algorithm from a restricted class of TLA$^+$ specifications into the input language of Cubicle. A prototypical implementation demonstrates the feasibility of the approach, although more work will be necessary to widen the scope of the translation. This work will be continued within the PARDI project, described in section 9.1 .

[0] http://cubicle.lri.fr

CARTE Team

7. New Results

7.1. Quantum Computing

Participants: Emmanuel Jeandel, Simon Perdrix, Renaud Vilmart.

- **ZX-calculus**

  The ZX-Calculus is a powerful graphical language for quantum mechanics and quantum information processing. The completeness of the language – i.e. the ability to derive any true equation – is a crucial question. In the quest for a complete ZX-calculus, supplementarity has been recently proved to be necessary for quantum diagram reasoning [73]. Roughly speaking, supplementarity consists in merging two subdiagrams when they are parameterized by antipodal angles. In [22], we introduce a generalised supplementarity – called cyclotomic supplementarity – which consists in merging n subdiagrams at once, when the n angles divide the circle into equal parts. We show that when n is an odd prime number, the cyclotomic supplementarity cannot be derived, leading to a countable family of new axioms for diagrammatic quantum reasoning. We exhibit another new simple axiom that cannot be derived from the existing rules of the ZX-Calculus, implying in particular the incompleteness of the language for the so-called Clifford+T quantum mechanics. We end up with a new axiomatisation of an extended ZX-Calculus, including an axiom schema for the cyclotomic supplementarity. This work has been presented at MFCS 2017 [22].

  The ZX-Calculus is devoted to represent complex quantum evolutions. But the advantages of quantum computing still exist when working with rebits, and evolutions with real coefficients. Some models explicitly use rebits, but the ZX-Calculus cannot handle these evolutions as it is. Hence, in [21], we define an alternative language solely dealing with real matrices, with a new set of rules. We show that three of its non-trivial rules are not derivable from the other ones and we prove that the language is complete for the $\pi/2$-fragment. We define a generalisation of the Hadamard node, and exhibit two interpretations from and to the ZX-Calculus, showing the consistency between the two languages. This work has been presented at QPL 2017 [21].

- **Causality and Quantum Computing**

  Since the classic no-go theorems by [43] and [65], contextuality has gained great importance in the development of quantum information and computation. This key characteristic feature of quantum mechanics represents one of the most valuable resources at our disposal to break through the limits of classical computation and information processing, with various concrete application

  An important class of contextuality arguments in quantum foundations are the All-versus-Nothing (AvN) proofs, generalising a construction originally due to Mermin. In [11], we present a general formulation of All-versus-Nothing arguments, and a complete characterisation of all such arguments which arise from stabiliser states. We show that every AvN argument for an n-qubit stabiliser state can be reduced to an AvN proof for a three-qubit state which is local Clifford-equivalent to the tripartite GHZ state. This is achieved through a combinatorial characterisation of AvN arguments, the AvN triple Theorem, whose proof makes use of the theory of graph states. This result enables the development of a computational method to generate all the AvN arguments in Z2 on n-qubit stabiliser states. We also present new insights into the stabiliser formalism and its connections with logic. This work has been presented at QPL 2017 [25] and published in the Philosophical Transactions of the Royal Society A [11].

Analyzing pseudo-telepathy graph games, we propose in [15] a way to build contextuality scenarios exhibiting the quantum supremacy using graph states. We consider the combinatorial structures generating equivalent scenarios. We investigate which scenarios are more multipartite and show that there exist graphs generating scenarios with a linear multipartiteness width. This work has been presented at FCT 2017 [15].

- **Measurement-based Quantum Computing**

  Measurement-based quantum computing (MBQC) is a universal model for quantum computation [74]. The combinatorial characterisation of determinism in this model [51], [48], [69], powered by measurements, and hence, fundamentally probabilistic, is the cornerstone of most of the breakthrough results in this field. The most general known sufficient condition for a deterministic MBQC to be driven is that the underlying graph of the computation has a particular kind of flow called Pauli flow. The necessity of the Pauli flow was an open question [48]. In [23], we show that the Pauli flow is necessary for real-MBQC, and not in general providing counterexamples for (complex) MBQC. We explore the consequences of this result for real MBQC and its applications. Real MBQC and more generally real quantum computing is known to be universal for quantum computing. Real MBQC has been used for interactive proofs by McKague. The two-prover case corresponds to real-MBQC on bipartite graphs. While (complex) MBQC on bipartite graphs are universal, the universality of real MBQC on bipartite graphs was an open question. We show that real bipartite MBQC is not universal proving that all measurements of real bipartite MBQC can be parallelised leading to constant depth computations. As a consequence, McKague techniques cannot lead to two-prover interactive proofs. This work has been presented at FCT 2017 [23].

7.2. Cellular automata as a model of computation

Participants: Nazim Fatès, Irène Marcovici.

The reversibility of classical cellular automata (CA) was examined for the case where the updates of the system are random. In this context, with B. Sethi and S. Das (India), we studied a particular form of reversibility: the possibility of returning infinitely often to the initial condition after a random number of time step,s this is the recurrence property of the system. We analysed this property for the simple rules and described the communication graph of the system [33].

We studied how to coordinate a team of agents to locate a hidden source on a two-dimensional discrete grid. The challenge is to find the position of the source with only sporadic detections. This problem arises in various situations, for instance when insects emit pheromones to attract their partners. A search mechanism named infotaxis was proposed to explain how agents may progressively approach the source by using only intermittent detections. With Q. Ladeveze, an intern, we re-examined in detail the properties of our bio-inspired algorithm that relies on the Reaction–Diffusion–Chemotaxis aggregation scheme to group agents that have limited abilities [38].

To study the robustness of asynchronous CA, we examined the coalescence phenomenon, which consists in observing the cases where two different initial conditions with the same sequence of updates quickly evolve to the same non-trivial configuration. With J. Francès de Mas, an intern, we studied the rules which always coalesce and those which exhibit a phase transition between a coalescing and non-coalescing behaviour. We proposed some formal explanations of non-trivial rapid coalescence giving lower bounds for the coalescence time of ECA 154 and ECA 62, and some first steps towards finding their upper bounds in order to prove that they have, respectively, quadratic and linear coalescence time [34].

We studied random mixtures of two deterministic Elementary Cellular Automata. There are 8088 such rules, called, diploid cellular automata. We used numerical simulations to perform some steps in the exploration of this space. As the mathematical analysis of such systems is a difficult task, we used numerical simulations to get insights into the dynamics of this class of stochastic cellular automata. We examined phase transitions and various types of symmetry breaking [17].

7.3. Extension of computable functions

Participant: Mathieu Hoyrup.

We worked on the computable aspects of an elementary problem in real analysis: extending a continuous function on a larger domain. More precisely, if a real-valued function f is defined on an interval $[0, a)$ (with $0 < a < 1$) and is computable there, under which conditions can it be extended to a computable function on $[0, 1]$? Although this question has a very simple formulation, it does not have a simple answer. We obtained many results showing how the answer depends on a and on the way f converges at a. Surprisingly, this problem provides new characterizations of already existing classes of real numbers previously defined in computability theory. This work is joint with Walid Gomaa and has been presented at LICS 2017 [19].

7.4. Genericity of weakly computable objects

Participant: Mathieu Hoyrup.

Computability theory abounds with classes of objects, defined for instance in terms of the computability content of the objects. A natural problem is then to compare these classes and separate them when possible. In order to separate two classes, one has to build an object that belongs to one class but not the other. So this object has to be computable in one sense but not the other. We show that in many cases these computability properties have a topological interpretation, and that the object to build must be at the same time computable in some weak topology (*weakly computable*) but *generic* in a stronger topology. We prove a general theorem stating the existence of such objects, thus providing a very handy tool to separate many classes. We use it in the study of the extension of computable functions (previous result) and in other situations. These results are presented in [13].

CIDRE Project-Team

7. New Results

7.1. Intrusion Detection

7.1.1. Intrusion Detection in Distributed Systems

Alert Correlation: In large systems, multiple (host and network) Intrusion Detection Systems (IDS) and many sensors are usually deployed. They continuously and independently generate notifications (event's observations, warnings and alerts). To cope with this amount of collected data, alert correlation systems have to be designed. An alert correlation system aims at exploiting the known relationships between some elements that appear in the flow of low level notifications to generate high semantic meta-alerts. The main goal is to reduce the number of alerts returned to the security administrator and to allow a higher level analysis of the situation. However, producing correlation rules is a highly difficult operation, as it requires both the knowledge of an attacker, and the knowledge of the functionalities of all IDSes involved in the detection process. In the context of the PhD of Erwan Godefroy, we focus on the transformation process that allows to translate the description of a complex attack scenario into correlation rules and its assessment. We show that, once a human expert has provided an action tree derived from an attack tree, a fully automated transformation process can generate exhaustive correlation rules that would be tedious and error prone to enumerate by hand. This is a top-down approach to correlation rule generation. With the PhD of Charles Xosanavongsa, we tackle the problem of a bottom-up approach that consists in discovering automatically the events or alerts that have been produced by the attacker activity. The objective is to classify automatically all suspicious entries in heterogeneous logs relative to a given attack. This requires to exhibit all log entries that are causally linked, and permits to produce a correlation rule that could detect later a new occurence of the attack.

Intrusion Detection in Cloud Infrastructure: Prior to detecting intrusion, it can be useful to know how the supervised system is vulnerable to attacks. Such result is obtained during a risk analysis phase in usual systems. In the PhD thesis of Pernelle Mensah, we try to automate the generation of the description of all possible attacks against a Cloud infrastructure. This work is divided in two separate steps: (1) We first discover the topology of the virtual machines executing in the cloud infrastructure [16], [17] and (2) Build in a second phase a topological attack graph that represents all possible known attacks on the virtual infrastructure. This graph will be later used either to adapt counter-measures to known attacks, or to generate automatically correlation rules to detect the described attacks.

Inferring the normal behavior of an application: We propose an approach to detect intrusions that affect the behavior of distributed applications. To determine whether an observed behavior is normal or not (occurence of an attack), we rely on a model of normal behavior. This model has been built during an initial training phase (machine learning approach). During this preliminary phase, the application is executed several times in a safe environment. The gathered traces (sequences of actions) are used to generate an automaton that characterizes all these acceptable behaviors. To reduce the size of the automaton and to be able to accept more general behaviors that are close to the observed traces, the automaton is transformed. These transformations may lead to introduce unacceptable behaviors. Our current work solves this problem by characterizing the acceptable behaviors with invariant properties that they must verify. During the PhD thesis of David Lanoe, we enhanced the model building. Moreover, we assess this solution, by applying it to a distributed file system called XtreemFS. We show that it is possible to build the model of this given application, and to detect attack against XtreemFS, without producing too much false positives.

This approach is particularly appealing to detect intrusions in industrial control systems since these systems exhibit well-defined behaviors at different levels: network level (network communication patterns, protocol specifications, etc.), control level (continue and discrete process control laws), or even the state of the local resources (memory or CPU). Industrial control systems (ICS) can be subject to highly sophisticated attacks which may lead the process towards critical states. Due to the particular context of ICS, protection mechanisms

are not always practical, nor sufficient. On the other hand, developing a process-aware intrusion detection solution with satisfactory alert characterization remains an open problem. Sophisticated process-aware attacks targeting industrial control systems require adequate detection measures taking into account the physical process. We propose an approach relying on automatically mined process specifications to detect attacks on sequential control systems. The specifications are synthesized as monitors that read the execution traces and report violations to the operator. In contrast to other approaches, a central aspect of our method consists in reducing the number of mined specifications suffering from redundancies. We evaluate our approach on a hardware-in-the-loop testbed with a complex physical process model and discuss our approach's mining efficiency and attack detection capabilities. This work has been submitted to the Safeprocess'18 conference.

7.1.2. Illegal Information Flow Detection

Our research work on intrusion detection based on information flow has been initiated in 2002. This research work has resulted in Blare, a framework for Intrusion Detection Systems [0], including KBlare, an implementation as a Linux Security Module (LSM), JBlare, an implementation for the Java Virtual Machine (JVM), and AndroBlare, for Android applications.

Information Leaks: Qualitative information flow aims at detecting information leaks, whereas the emerging quantitative techniques target the estimation of information leaks. Quantifying information flow in the presence of low inputs is challenging, since the traditional techniques of approximating and counting the reachable states of a program no longer suffice. We propose an automated quantitative information flow analysis for imperative deterministic programs with low inputs. The approach relies on a novel abstract domain, the cardinal abstraction, in order to compute a precise upper-bound over the maximum leakage of batch-job programs. We prove the soundness of the cardinal abstract domain by relying on the framework of abstract interpretation. We also prove its precision with respect to a flow-sensitive type system for the two-point security lattice. This approach has been published in POPL'17 [8].

Correct information flow monitoring by design: As mentionned previously, our research team is developing an information monitor called Blare. Like most of its competitors (e.g. Laminar or Weir) our solution is based on the Linux Security Module (LSM) framework. However, this framework was initially designed with access control in mind. A natural question arises from this matter of fact: does the LSM framework can be used to correctly track information flow (at the operating system level) ? In the context of his PhD thesis, Laurent Georget has studied this very same question.

To tackle this problem, Laurent Georget has designed an ad hoc static analysis that run as a GCC plugin during the Linux kernel compilation. This analysis can prove (or disprove) the fact that LSM hooks within a chosen set of system calls (known to realize information flows between operating systems containers like files, sockets or pipe) are placed at correct locations so as to intercept these possible information flows. The experiments conducted by Laurent Georget have revealed that on an initial set of 38 system calls, 28 were correctly instrumented by LSM, 4 of them were equipped with a LSM hook that could miss some information flow (under certain circumstances), 3 were simply lacking a LSM hook, and 3 false positives had to be manually analyzed and requalified. Laurent Georget was able to produce a kernel patch to remove all missing and misplaced hooks. This patch can be prove to be correct using the same tool. This contribution was published at FormaliSE 2017 [12].

We had detected for a long time a subtle bug in our information flow monitor implementation (Blare) that we were able to track down to a race condition between two concurrent system calls reading and writing into the same pipe. Laurent Georget has proposed during its PhD an elegant solution to this complex problem: he proposed to divide each information flow into three stages: the activation, the execution and the deactivation. Only the activation and deactivation can be observed by the monitor using LSM hooks placed at the beginning and the exit of a system call. This way, it becomes possible to track causal dependencies between concurrent system calls within the LSM framework. Laurent Georget has proved (using the Coq proof assistant) that his approach is correct and computes the smallest possible over-approximation, in the sense that for any concurrent execution where multiple system calls are used there exists a linearization of this execution that produces the

[0] http://www.blare-ids.org/

information flow computed by his algorithm. Laurent Georget has implemented his algorithm in the Linux kernel. This contribution was publish at Software Engineering & Formal Methods (2017) where it was granted the best paper award [11]. Laurent Georget has defended his PhD thesis in September 2017 .

Advanced Persistent Threats: Long lived attack campaigns known as Advanced Persistent Threats (APTs) have emerged as a serious security risk. These attack campaigns are customised for their target and performed step by step during months on end. The major difficulty in detecting an APT is keeping track of the different steps logged over months of monitoring and linking them. In [29], we described TerminAPTor, an APT detector which highlights links between the traces left by attackers in the monitored system during the different stages of an attack campaign. TerminAPTor tackles this challenge by resorting to Information Flow Tracking (IFT). TerminAPTor was presented last year and we have pursue our effort in this area. More precisely, we have focus on the evaluation of this solution and thus we face to the lack of public datasets of attacks. We develop Moirai a framework dedicated to attacks scenario sharing [22] .

Characterizing Android Malware: Android has become the world's most popular mobile operating system, and consequently the most popular target for unscrupulous developers. These developers seek to make money by taking advantage of Android users who customize their devices with various applications, which are the main malware infection vector. Indeed, the most likely way a user executes a repackaged application is by downloading a seemingly harmless application from a store and executing it. Such an application may have been modified by an attacker in order to add malicious pieces of code.

To fight repackaged applications containing malicious code, most official application marketplaces have implemented security analysis tools that try to detect and remove malware. Countermeasures adopted by the attackers to bypass these new controls can be divided into two main approaches: avoiding static analysis and avoiding dynamic analysis. A static analysis of an application consists of analysing its code and its resources without executing it. Conversely, dynamic analysis stands for any kind of analysis that requires executing the application in order to observe its actions.

The Kharon project [30] goes a step further from classical dynamic analysis of malware [0]. Funded by the Labex CominLabs and involving partners of CentraleSupélec, Inria and INSA Centre Val de Loire, this project aims to capture a compact and comprehensive representation of malware. To achieve such a goal we have developed tools to monitor operating systems' information flows induced by the execution of a marked application. We support the idea that the best way to understand malware impact is to observe it in its normal execution environment i.e., a real smartphone. Additionally, the main challenge is to be able to trigger malicious behaviors even when the malware tries to escape dynamic analysis.

In this context, we have developed an original solution whose main purpose is a relavant dynamic analysis of the malicious code. We develop the GroddDroid software, that mainly consists of 'helping the malware to execute'. To reach this goal, GroddDroid relies on a previous static analysis that evidence all the execution paths leading to the malicious code. We compute a global control flow graph (CFG) that exhibits execution paths to reach specific parts of code, even if these paths use callbacks that are handled in the Android framework itself [15]. Finally, GroddDroid slightly modifies the bytecode of the infected application in order to defeat the protection against dynamic analysis and executes the suspicious code in its most favorable execution conditions. Thus, GroddDroid helps to understanding malware's objectives and the consequences on the health of a user's device.

GroddDroid can also be used for classifying applications between goodware and malware. We show in [19] that benign applications have a System Flow Graph (a graph that represent flows at operating system level) that can be anticipated. Malware that perform complex operations such as installing backdoor or launching a Tor client, have a CFG that differ enough to be classified easily.

Our main research direction and challenges in this area are to continue to enhance these technologies in order to reach a sufficient level of software maturity to deploy a permanent platform of malware analysis in the LHS (Laboratory of High Security) and to create new opportunities with industrial partners.

[0] http://kharon.gforge.inria.fr

7.1.3. Intrusion Detection in Low-Level Software Components

In order to protect the IDS itself, we have initiated different research activities in the domain of hardware security. Our goal is to use co-design software/hardware approaches against traditional software attacks. In a bilateral research project with HP Inc Research Labs, we investigate how dedicated hardware could be used to monitor the whole software stack (from the firmware to the user-mode applications). In the CominLabs HardBlare project, we study the use of a dedicated co-processor to enforce Information Flow Control (IFC) on the main CPU. Finally, in the context of the PhD thesis of Thomas Letan (ANSSI), we investigate the use of formal methods to evaluate the security guarantees provided by hardware platforms, which combine different CPUs, chipsets and memories.

Highly privileged software, such as firmware, is an attractive target for an attacker. Thus, BIOS vendors use cryptographic signatures to ensure firmware integrity at boot time. Nevertheless, such boot time protection does not prevent an attacker from exploiting vulnerabilities at runtime. To detect such runtime attacks, we proposed an event-based monitoring approach that relies on an isolated co-processor [10]. We instrument the code executed on the main CPU to send information about its behavior to the monitor. In this work, we focus on the detection of attacks targeting the System Management Mode (SMM), a highly privileged x86 execution mode executing firmware code at runtime. We use the control flow of the code as a model of its behavior. We evaluate our approach with two open-source implementations: EDK II and coreboot. We evaluate its ability to detect state-of-the-art attacks and its runtime execution overhead by simulating an x86 system coupled with an ARM Cortex A5 co-processor. The results show that our solution detects intrusions from the state of the art while remaining acceptable in terms of performance overhead in the context of the SMM. This work has been done in collaboration with HP Inc Research Labs, in the context of the PhD of Ronny Chevalier.

Over time, hardware designs have constantly grown in complexity and modern platforms involve multiple interconnected hardware components. During the last decade, several vulnerability disclosures have proven that trust in hardware can be misplaced. The approach we developed with Thomas Letan rely on a formal definition of Hardware-based Security Enforcement (HSE) mechanisms, a class of security enforcement mechanisms such that a software component relies on the underlying hardware platform to enforce a security policy. We then model a subset of a x86-based hardware platform specifications and we prove the soundness of a realistic HSE mechanism within this model using Coq, a proof assistant system.

The HardBlare project proposes a software/hardware co-design methodology to ensure that security properties are preserved all along the execution of the system but also during files storage. It is based on the Dynamic Information Flow Tracking (DIFT) that generally consists in attaching tags to denote the type of information that are saved or generated within the system. These tags are then propagated when the system evolves and information flow control is performed in order to guarantee the safe execution and storage within the system monitored by security policies. We proposed ARMHEx [20], a practical solution targeting DIFT on ARM-based SoCs (e.g. Xilinx Zynq). Current DIFT implementations suffer from two major drawbacks. First, recovering required information for DIFT is generally based on software instrumentation leading to high time overheads. ARMHEx takes profit of ARM CoreSight debug components and static analysis to drastically reduce instrumentation time overhead (up to 90% compared to existing works). Then, security of the DIFT hardware extension itself is not considered in related works. In this work, we tackle this issue by proposing a solution based on ARM Trustzone. This work has been done in the context of the PhD of Muhammad Abdul Wahab and Mounir Nasr Allah.

7.1.4. Vizualization

When using Intrusion Detection Systems (IDS), the large quantities of alerts generated are difficult to handle by security experts. To help solving this problem, we have proposed VEGAS, an alerts visualization and classification tool that allows primary visions based on their principal component analysis (PCA) representation. Following this, we have studied the context of collaboration between the various security actors. We have then proposed an extension to VEGAS that allows to help the actors to collaborate. We have developped an interface that permits the front-end operator to quickly understand the security events, and group them to organize incidents and send them to dedicated analysts. Conversely, once the incidents have

been analysed, the analysts can send information to the front-line operators to help them understanding the futur security events.

We also developed another tool called STARLORD [14] that permits to an administrator the explore in a 3D graph representing the links between the heterogeneous entries in various logs produced either by the system, applications or IDSes. To emphasize the important relations between the lines of logs that can potentially be part of an attack activity, we classify these links in order to present only the part of the graph that is linked to an indicators of compromission.

Our previous research on visualization of security events has lead to two proofs-of-concept (See ELVIS and CORGI softwares). We are currently pursuing business opportunities on this topic. Indeed SplitSec is a soon to be founded startup developing tools to help security experts to better manage and understand security data. Scalable analysis solutions and data visualisations adapted for security are combined into powerful tools for incident response. Until June 2017, Christopher Humphries has been hired by Inria as a technology transfer engineer to build these tools based on promising research prototypes.

7.2. Privacy

7.2.1. Image Encryption

More and more users prefer to share their photos through image-sharing platforms of social networks than using e-mail or personal webpages. Since the provider of the image-sharing platform can clearly know the contents of any published images, the users have to trust the provider to respect their privacy or has to encrypt their images. In the context of the PhD of Kun He, we have proposed an IND-CPA image encryption algorithm that preserve the image format after encryption, and we have shown that our encryption algorithm can be used on several widely used image-sharing platforms such as Flickr, Pinterest, Google+ and Twitter. Kun He has completed her PhD thesis in September 2017 [5].

7.3. Security of Communicating and Distributed Systems

7.3.1. Routing Protocol for Tactical Mobile Ad Hoc Networks

In the context of the PhD thesis of Florian Grandhomme, we propose new secure and efficient algorithms and protocols to provide inter-domain routing in the context of tactical mobile ad hoc network. The proposed protocol has to handle context modification due to the mobility of Mobile Ad hoc NETwork (MANET), that is to say split of a MANET, merge of two or more MANET, and also handle heterogeneity of technology and infrastructure. The solution has to be independent from the underlying intra-domain routing protocol and from the infrastructure: wired or wireless, fixed or mobile. This work is done in cooperation with DGA-MI.

New generation military equipment, soldiers and vehicles, use wireless technology to communicate on the battlefield. During missions, they form a MANET. Since the battlefield includes coalition, each group may communicate with another group, and inter-MANET communication may be established. Inter-MANET (or inter-domain MANET) communication should allow communication, but maintain a control on the exchanged information. Several protocols have been proposed in order to handle inter-domain routing for tactical MANETs. During the thesis we have shown that simulator (NS3) or emulator (CORE) do not handle correctly ad hoc network behavior and then that solution in the state of the art are more complex than needed. Based on this analysis, we propose some preconizations to design Inter-domain protocols for MANET and we propose the ITMAN (Inter Tactical Mobile Ad hoc Network) protocol that allows also to handle simple routing policy (merge, link and deny). We evaluate this new protocol through experimentation and we show that our proposition is quite efficient. On going work on this protocol is the definition and implementation of more subtle routing policy that allow announce filtering of giving prefix for example.

7.3.2. Decentralized Cryptocurrency Systems

Distributed Ledgers (e.g. Bitcoin) occupy currently the first lines of the economical and political media and many speculations are done with respect to their level of coherence and their computability power. Interestingly, there is no consensus on the properties and abstractions that fully capture the behaviour of distributed ledgers. The interest in formalising the behaviour of distributed ledgers is twofold. Firstly, it helps to prove the correctness of the algorithms that implement existing distributed ledgers and explore their limits with respect to an unfriendly environment and target applications. Secondly, it facilitates the identification of the minimal building blocks necessary to implement the distributed ledger in a specific environment. Even though the behaviour of distributed ledgers is similar to abstractions that have been deeply studied for decades in distributed systems no abstraction is sufficiently powerful to capture the distributed ledger behaviour. We have defined the Distributed Ledger Register, a register that mimics the behaviour of one of the most popular distributed ledger, i.e. the Bitcoin ledger. The aim of our work is to provide formal guarantees on the coherent evolution of Bitcoin. We furthermore showed that the Bitcoin blockchain maintenance algorithm verifies the distributed ledger register properties under strict conditions. Moreover, we proved that the Distributed Ledger Register verifies the regularity register specification. It follows that the strongest coherency implemented by Bitcoin is regularity under strong assumptions (i.e. partial synchronous systems and sparse reads). In [7] we proposed a study that contradicts the common belief that Bitcoin implements strong coherency criteria in a totally asynchronous system. To the best of our knowledge, our work is the first one that makes the connection between the distributed ledgers and the classical theory of distributed shared registers.

Double spending and blockchain forks are two main issues that the Bitcoin crypto-system is confronted with. The former refers to an adversary's ability to use the very same coin more than once while the latter reflects the occurrence of transient inconsistencies in the history of the blockchain distributed data structure. We present a new approach to tackle these issues: it consists in adding some local synchronization constraints on Bitcoin's validation operations, and in making these constraints independent from the native blockchain protocol. Synchronization constraints are handled by nodes which are randomly and dynamically chosen in the Bitcoin system. In [13] we show that with such an approach, content of the blockchain is consistent with all validated transactions and blocks which guarantees the absence of both double-spending attacks and blockchain forks.

7.3.3. Large Scale Systems

Population Protocol: the computational model of population protocols is a formalism that allows the analysis of properties emerging from simple and pairwise interactions among a very large number of anonymous finite-state agents. Significant work has been done so far to determine which problems are solvable in this model and at which cost in terms of states used by the protocols and time needed to converge. The problem tackled in is the population proportion problem: each agent starts independently from each other in one of two states, say A or B, and the objective is for each agent to determine the proportion of agents that initially started in state A, assuming that each agent only uses a finite set of state, and does not know the number n of agents. In [18], we show that for any $\delta \in (0,1)$, the number of interactions needed per node to converge is $O(ln(n/\delta))$ with probability at least $1-\delta$. We also prove that each node can determine, with any high probability, the proportion of nodes that initially started in a given state without knowing the number of nodes in the system. This work provides a precise analysis of the convergence bounds, and shows that using the 4-norm is very effective to derive useful bounds.

Distributed Stream Processing Systems: shuffle grouping is a technique used by stream processing frameworks to share input load among parallel instances of stateless operators. With shuffle grouping each tuple of a stream can be assigned to any available operator instance, independently from any previous assignment. A common approach to implement shuffle grouping is to adopt a Round-Robin policy, a simple solution that fares well as long as the tuple execution time is almost the same for all the tuples. However, such an assumption rarely holds in real cases where execution time strongly depends on tuple content. As a consequence, parallel stateless operators within stream processing applications may experience unpredictable unbalance that, in the end, causes undesirable increase in tuple completion times. In [25] we propose Online Shuffle Grouping

(OSG), a novel approach to shuffle grouping aimed at reducing the overall tuple completion time. OSG estimates the execution time of each tuple, enabling a proactive and online scheduling of input load to the target operator instances. Sketches are used to efficiently store the otherwise large amount of information required to schedule incoming load. We provide a probabilistic analysis and illustrate, through both simulations and a running prototype, its impact on stream processing applications.

The real time analysis of massive data streams is of utmost importance in data intensive applications that need to detect as fast as possible and as efficiently as possible (in terms of computation and memory space) any correlation between its inputs or any deviance from some expected nominal behavior. The IoT infrastructure can be used for monitoring any events or changes in structural conditions that can compromise safety and increase risk. It is thus a recurrent and crucial issue to determine whether huge data streams, received at monitored devices, are correlated or not as it may reveal the presence of attacks. We propose a metric, called codeviation, that allows to evaluate the correlation between distributed massive streams. This metric is inspired from classical metric in statistics and probability theory, and as such enables to understand how observed quantities change together, and in which proportion. In [6], we propose to estimate the codeviation in the data stream model. In this model, functions are estimated on a huge sequence of data items, in an online fashion, and with a very small amount of memory with respect to both the size of the input stream and the values domain from which data items are drawn. We then generalize our approach by presenting a new metric, the Sketch-metric, which allows us to define a distance between updatable summaries of large data streams. An important feature of the Sketch-metric is that, given a measure on the entire initial data streams, the Sketch-metric preserves the axioms of the latter measure on the sketch. We finally conducted extensive experiments on both synthetic traces and real data sets allowing us to validate the robustness and accuracy of our metrics.

COMETE Project-Team

6. New Results

6.1. Foundations of information hiding

Information hiding refers to the problem of protecting private information while performing certain tasks or interactions, and trying to avoid that an adversary can infer such information. This is one of the main areas of research in Comète; we are exploring several topics, described below.

6.1.1. Information Leakage Games

In [19] we studied a game-theoretic setting to model the interplay between attacker and defender in the context of information flow, and to reason about their optimal strategies. In contrast with standard game theory, in our games the utility of a mixed strategy is a convex function of the distribution on the defender's pure actions, rather than the expected value of their utilities. Nevertheless, the important properties of game theory, notably the existence of a Nash equilibrium, still hold for our (zero-sum) leakage games, and we provided algorithms to compute the corresponding optimal strategies. As typical in (simultaneous) game theory, the optimal strategy is usually mixed, i.e., probabilistic, for both the attacker and the defender. From the point of view of information flow, this was to be expected in the case of the defender, since it is well known that randomization at the level of the system design may help to reduce information leaks. Regarding the attacker, however, this seems the first work (w.r.t. the literature in information flow) proving formally that in certain cases the optimal attack strategy is necessarily probabilistic.

6.1.2. Efficient Utility Improvement for Location Privacy

The continuously increasing use of location-based services poses an important threat to the privacy of users. A natural defense is to employ an obfuscation mechanism, such as those providing geo-indistinguishability [24], a framework for obtaining formal privacy guarantees that has become popular in recent years. Ideally, one would like to employ an optimal obfuscation mechanism, providing the best utility among those satisfying the required privacy level. In theory optimal mechanisms can be constructed via linear programming. In practice, however, this is only feasible for a radically small number of locations. As a consequence, all known applications of geo-indistinguishability simply use noise drawn from a planar Laplace distribution.

In [12], we studied methods for substantially improving the utility of location obfuscation, while maintaining practical applicability as a main goal. We provided such solutions for both infinite (continuous or discrete) as well as large but finite domains of locations, using a Bayesian remapping procedure as a key ingredient. We evaluated our techniques in two real world complete datasets, without any restriction on the evaluation area, and showed important utility improvements with respect to the standard planar Laplace approach.

6.1.3. Trading Optimality for Performance in Location Privacy

Location-Based Services (LBSs) provide invaluable aid in the everyday activities of many individuals, however they also pose serious threats to the user' privacy. There is, therefore, a growing interest in the development of mechanisms to protect location privacy during the use of LBSs. Nowadays, the most popular methods are probabilistic, and the so-called optimal method achieves an optimal trade-off between privacy and utility by using linear optimization techniques.

Unfortunately, due to the complexity of linear programming, the method is unfeasible for a large number N of locations, because the constraints are $O(N^3)$. In [20], we have proposed a technique to reduce the number of constraints to $O(N^2)$, at the price of renouncing to perfect optimality. We have showed however that on practical situations the utility loss is quite acceptable, while the gain in performance is significant.

6.1.4. Methods for Location Privacy: A comparative overview

The growing popularity of location-based services, allowing to collect huge amounts of information regarding users' location, has started raising serious privacy concerns. In [13] we analyzed the various kinds of privacy breaches that may arise in connection with the use of location-based services, and we surveyd and compared the metrics and the mechanisms that have been proposed in the literature.

6.1.5. Quantifying Leakage in the Presence of Unreliable Sources of Information

Belief and min-entropy leakage are two well-known approaches to quantify information flow in security systems. Both concepts stand as alternatives to the traditional approaches founded on Shannon entropy and mutual information, which were shown to provide inadequate security guarantees. In [16] we unified the two concepts in one model so as to cope with the frequent (potentially inaccurate, misleading or outdated) attackers' side information about individuals on social networks, online forums, blogs and other forms of online communication and information sharing. To this end we proposed a new metric based on min-entropy that takes into account the adversary's beliefs.

6.1.6. Differential Inference Testing: A Practical Approach to Evaluate Anonymized Data

In order to protect individuals' privacy, governments and institutions impose some obligations on data sharing and publishing. Mainly, they require the data to be "anonymized". In this paper, we have shortly discussed the criteria introduced by European General Data Protection Regulation to assess anonymized data. We have argued that the evaluation of anonymized data should be based on whether the data allows individual based inferences, instead of being centered around the concept of re-identification as the regulation has proposed.

Then, we have proposed an inference-based framework that can be used to evaluate the robustness of a given anonymized dataset against a specific inference model, e.g. a machine learning model.

Our approach evaluates the anonymized data itself, and deals with the related anonymization technique as a black-box. Thus, it can be used to assess datasets that are anonymized by organizations which may prefer not to provide access to their techniques. Finally, we have used our framework to evaluate two datasets after being anonymized using k-anonymity and l-diversity.

6.1.7. Formal Analysis and Offline Monitoring of Electronic Exams

More and more universities are moving toward electronic exams (in short e-exams). This migration exposes exams to additional threats, which may come from the use of the information and communication technology. In [17], we have identified and defined several security properties for e-exam systems. Then, we have showed how to use these properties in two complementary approaches: model-checking and monitoring.

We have illustrated the validity of our definitions by analyzing a real e-exam used at the pharmacy faculty of University Grenoble Alpes (UGA) to assess students. On the one hand, we have instantiated our properties as queries for ProVerif, a process calculus based automatic verifier for cryptographic protocols,

and we have used it to check our modeling of UGA exam specifications. ProVerif found some attacks. On the other hand, we have expressed our properties as Quantified Event Automata (QEAs), and we have synthesized them into monitors using MarQ, a Java tool designed to implement QEAs. Then, we have used these monitors to verify real exam executions conducted by UGA. Our monitors found fraudulent students and discrepancies between the specifications of UGA exam and its implementation.

6.1.8. On the Compositionality of Quantitative Information Flow

In the min-entropy approach to quantitative information flow, the leakage is defined in terms of a minimization problem, which, in the case of large systems, can be computationally rather heavy. The same happens for the recently proposed generalization called g-vulnerability. In [18] we studied the case in which the channel associated to the system can be decomposed into simpler channels, which typically happens when the observables consist of several components. Our main contribution is the derivation of bounds on the g-leakage of the whole system in terms of the g-leakages of its components. We also considered the particular cases of min-entropy leakage and of parallel channels, generalizing and systematizing results from the literature. We demonstrated the effectiveness of our method and evaluate the precision of our bounds using examples.

6.2. Foundations of Concurrency

Distributed systems have changed substantially in the recent past with the advent of phenomena like social networks and cloud computing. In the previous incarnation of distributed computing the emphasis was on consistency, fault tolerance, resource management and related topics; these were all characterized by *interaction between processes*. Research proceeded along two lines: the algorithmic side which dominated the Principles Of Distributed Computing conferences and the more process algebraic approach epitomized by CONCUR where the emphasis was on developing compositional reasoning principles. What marks the new era of distributed systems is an emphasis on managing access to information to a much greater degree than before.

6.2.1. Declarative Framework for Semantical Interpretations of Structured Information — An Applicative Approach.

Spatial constraint systems are algebraic structures from concurrent constraint programming to specify spatial and epistemic behavior in multi-agent system. In [21], [15] we studied the applicability of declarative models to encode and describe structured information by means of semantics. Specifically, we introduced D-SPACES, an implementation of constraint systems with space and extrusion operators. D-SPACES provides property-checking methods as well as an implementation of a specific type of constraint systems (a spatial boolean algebra). We showed the applicability of this framework with two examples; a scenario in the form of a social network where users post their beliefs and utter their opinions, and a semantical interpretation of a logical language to express time behaviors and properties.

6.2.2. Characterizing Right Inverses for Spatial Constraint Systems with Applications to Modal Logic

In [23] spatial constraint systems were used to give an abstract characterization of the notion of normality in modal logic and to derive right inverse/reverse operators for modal languages. In particular, a necessary and sufficient condition for the existence of right inverses was identified and the abstract notion of normality is shown to correspond to the preservation of finite suprema. Furthermore, a taxonomy of normal right inverses was provided, identifying the greatest normal right inverse as well as the complete family of minimal right inverses. These results were applied to existing modal languages such as the weakest normal modal logic, Hennessy-Milner logic, and linear-time temporal logic. Some implications of these results were also discussed in the context of modal concepts such as bisimilarity and inconsistency invariance.

6.2.3. Observational and Behavioural Equivalences for Soft Concurrent Constraint Programming

In citegadducci:hal-01675060 we presented a labelled semantics for Soft Concurrent Constraint Programming (SCCP), a meta-language where concurrent agents may synchronise on a shared store by either posting or checking the satisfaction of (soft) constraints. SCCP generalises the classical formalism by parametrising the constraint system over an order-enriched monoid, thus abstractly representing the store with an element of the monoid, and the standard unlabelled semantics just observes store updates. The novel operational rules were shown to offer a sound and complete co-inductive technique to prove the original equivalence over the unlabelled semantics. Based on this characterisation, we provided an axiomatisation for finite agents.

6.2.4. On the Expressiveness of Spatial Constraint Systems

The dissertation [11] focused on the expressiveness of spatial constraint systems in the broader perspective of modal and epistemic behaviour. It was shown that that spatial constraint systems are sufficiently robust to capture inverse modalities and to derive new results for modal logics. It was shown that one can use scs's to express a fundamental epistemic behaviour such as knowledge. The dissertation also provided an algebraic characterization of the notion of distributed information by means of constructors over scs's.

DATASPHERE Team

7. New Results

7.1. Political economy

We pursued our work on digital platforms and their impact on the structure of socio-economic systems, which results from the capacity to separate data or information from the actors of the physical world. In [14], we showed how the movement above ground of the intermediation activity transforms territories. We developed further this idea in [5] to suggest that a new political grammar is necessary to understand the relationships between the actors.

7.2. Anthropocene

In [6], we investigated the possible similarities between biological systems and social systems facing shortage of resources, suggesting that the digital revolution might have something to do with the Anthropocene. The timing of the digital revolution was further investigated in [7], to further analyse the relationships between the two transformation affecting the contemporary period. An investigation of the world of images and photography in the time of algorithms was conducted in [4].

7.3. Laws and digital

The emergence of digital services affects the legal system. The law is always associated to a territory, while digital systems act remotely over large regions crossing borders to reach the population, imposing new norms. In [3], [2], we suggest that a new framework is necessary to apprehend new phenomena, such as the those resulting from the conflicts between global search engines and local rules with respect to the Right to be forgotten for instance.

7.4. Network data analytics

In collaboration with the Chinese Academy of Sciences, we worked on packet processing algorithmic for high speed network measurements. In [9] a packet capture archive system is developed and described. In [8] a theoretical analysis of the TCAM updates delay that is the main shortcoming of TCAM usage in high speed packet processors is presented. Quality of service for network functions were considered in [13].

7.5. Data protection

We developed new mechanisms to process aggregation while preserving the confidentiality of the initial data in the framework of Robert Riemann's thesis [1]. The benefits of distributed protocols for online voting was considered [12]. A distributed aggregation mechanisms preserving confidentiality of data based on Kademlia was proposed in [11]. Applications of the previous algorithms to lotteries was considered in [10].

PESTO Project-Team

7. New Results

7.1. Modelling

7.1.1. New protocol and adversary models

Participants: Jannik Dreier, Steve Kremer, Ludovic Robin.

Symbolic models for security protocol verification, following the seminal ideas of Dolev and Yao, come in many flavors, even though they share the same ideas. A common assumption is that the attacker has complete control over the network: he can therefore intercept any message. Depending on the precise model this may be reflected either by the fact that any protocol output is directly routed to the adversary, or communications may be among any two participants, including the attacker—the scheduling between which exact parties the communication happens is left to the attacker. These two models may seem equivalent at first glance and, depending on the verification tools, either one or the other semantics is implemented. In collaboration with Babel (IIT Bombay) we show that, unsurprisingly, they indeed coincide for reachability properties. However, when we consider equivalence properties, we prove that these two semantics are incomparable. We also introduce a new semantics, where internal communications are allowed but messages are always eavesdropped by the attacker. We show that this new semantics yields strictly stronger equivalence relations and identify two subclasses of protocols for which the three semantics coincide. These results were presented at POST'17 [16].

Isolated Execution Environments (IEEs), such as ARM TrustZone and Intel SGX, offer the possibility to execute sensitive code in isolation from other, potentially malicious programs, running on the same machine, or a potentially corrupted OS. A key feature of IEEs is the ability to produce reports binding cryptographically a message to the program that produced it, typically ensuring that this message is the result of the given program running on an IEE. In collaboration with Jacomme (ENS Cachan) and Scerri (Univ Bristol), Kremer presented a symbolic model for specifying and verifying applications that make use of such features. For this they introduced the SℓAPiC process calculus to reason about reports issued at given locations. They also provide tool support, extending the *SAPIC/TAMARIN* toolchain and demonstrate the applicability of their framework on several examples implementing secure outsourced computation (SOC), a secure licensing protocol and a one-time password protocol that all rely on such IEEs. This work has been published and presented at EuroS&P'17 [30].

Modern security protocols may involve humans in order to compare or copy short strings between different devices. Multi-factor authentication protocols, such as Google 2-factor or 3D-secure are typical examples of such protocols. However, such short strings may be subject to brute force attacks. In collaboration with Delaune (IRISA), we propose a symbolic model which includes attacker capabilities for both guessing short strings, and producing collisions when short strings result from an application of weak hash functions. We propose a new decision procedure for analysing (a bounded number of sessions of) protocols that rely on short strings. The procedure has been integrated in the *Akiss* tool and tested on protocols from the ISO/IEC 9798-6:2010 standard. This work has been published and presented at CSF'17 [26].

Most security properties are modelled as *safety* properties (*"bad things do not happen"*). Another important class of properties is that of *liveness* properties (*"eventually, good things happen"*). Reasoning about the class of *liveness* properties of cryptographic protocols, has received little attention in the literature, even though this class is vital in many security-sensitive applications, such as fair exchange protocols, or security layers in industrial control systems. In collaboration with Backes and Künnemann (Univ Saarland, Germany), Dreier and Kremer have designed a protocol and adversary model that are suitable for reasoning about liveness properties. Tool support is also provided by extending the *SAPIC/TAMARIN* tool chain and several case studies demonstrate the effectiveness of the approach. This work has been published and presented at EuroS&P'17 [17].

7.1.2. New properties

Participant: Jannik Dreier.

Industrial systems are nowadays regularly the target of cyberattacks, the most famous being Stuxnet[0]. At the same time such systems are increasingly interconnected with other systems and insecure media such as Internet. In contrast to other IT systems, industrial systems often do not only require classical properties like data confidentiality or authentication of the communication, but have special needs due to their interaction with the physical world. For example, the reordering or deletion of some commands sent to a machine can cause the system to enter an unsafe state with potentially catastrophic effects. To prevent such attacks, the integrity of the message flow is necessary.

In joint work with Lafourcade (Univ Clermont-Ferrand), Potet, and Puys (Univ Grenoble Alpes), Dreier developed a formal definition of Flow Integrity in the context of industrial systems. The framework is applied to two well-known industrial protocols: OPC-UA and MODBUS. Using *TAMARIN*, they identified several design flaws in some of the different versions of these protocols. They also discussed how to efficiently model counters and timestamps in *TAMARIN*, as they are key ingredients of the analyzed protocols. This work was presented at SECRYPT'17 [32], and won a Best Student Paper Award.

7.2. Analysis

7.2.1. Analysis of equivalence properties

Participants: Vincent Cheval, Véronique Cortier, Antoine Dallon, Ivan Gazeau, Steve Kremer, Joseph Lallemand, Itsaka Rakotonirina, Christophe Ringeissen.

Automatic tools based on symbolic models have been successful in analyzing security protocols. These tools are particularly well adapted for trace properties (e.g. secrecy or authentication). However, they often fail to analyse equivalence properties. Equivalence properties can express a variety of security properties, including in particular privacy properties (vote privacy, anonymity, untraceability). Several decision procedures have already been proposed but the resulting tools are often rather limited, and lack efficiency.

In the case of a passive adversary, Ringeissen, in collaboration with Marshall (Univ Mary Washington, USA) and Erbatur (LMU, Germany) present new combination techniques for the study of deducibility and static equivalence in unions of equational theories sharing constructors. This allows us to develop new modularity results for the decidability of deducibility and static equivalence. In turn, this should allow for the security analysis of protocols which previous disjoint combination methods could not address because their axiomatization corresponds to the union of non-disjoint equational theories. This work has been presented at CADE'17 [28].

In case of an active adversary, and a bounded number of sessions, we made several advances. The *Akiss* tool has been extended in two directions. Gazeau and Kremer, in collaboration with Baelde (LSV, ENS Cachan) and Delaune (IRISA) have extended the underlying theory and the *Akiss* tool with support for exclusive or. They analyse unlinkability in several RFID protocols and resistance to guessing attacks of several password-based protocols. This work has been presented at CSF'17 [18]. Gazeau and Kremer also extended the *Akiss* tool to analyse protocols with else branches. This is particularly useful when verifying equivalence properties, as one needs to model precisely the error messages sent out when tests fail. While ignoring these branches may often be safe when studying trace properties this is not the case for equivalence properties, as for instance witnessed by an attack on the European electronic passport. One appealing feature of our approach is that our extension re-uses the saturation procedure which is at the heart of the verification procedure of *Akiss* as a black box, without need to modify it. As a result we obtain the first tool that is able verify equivalence properties for protocols that may use xor and else branches. We demonstrate the tool's effectiveness on several case studies, including the AKA protocol deployed in mobile telephony. This result was presented at ESORICS'17 [29]. Cortier and Dallon, in collaboration with Delaune (IRISA) propose a novel algorithm, based on graph planning and SAT-solving, which significantly improves the efficiency of the analysis of equivalence properties. The

[0] Stuxnet. https://en.wikipedia.org/wiki/Stuxnet

resulting implementation, SAT-Equiv, can analyze several sessions where most tools have to stop after one or two sessions. The approach has been presented at CSF'17 [20] for protocols with symmetric encryption and no else branches. Finally, Cheval, Kremer, and Rakotonirina have worked on complexity results for deciding equivalence properties and provide a decision procedure in the case of a bounded number of sessions. They showed that trace equivalence and labelled bisimilarity for a large variety of cryptographic primitives—those that can be represented by a subterm convergent destructor rewrite system— are both CoNEXP complete. Moreover, the procedure has been implemented in a new tool, *DeepSec*. Extensive experiments demonstrate that it is significantly more efficient than most other similar tools (being only slightly outperformed by SAT-Equiv in some specific examples), while at the same time raises the scope of the protocols that can be analysed. These results are currently under submission.

The previous results apply for a bounded number of sessions and may still be limited for a large number of sessions. In collaboration with Maffei and Grimm, Lallemand and Cortier have devised a novel approach [24] for proving equivalence properties. Instead of *deciding* equivalence, like for the previous approaches, they design a type system, sound w.r.t. equivalence. The resulting tool TypeEquiv can consider a bounded as well as an unbounded number of sessions, or a mix of both. It induces a significant speedup compared to previous tools for a bounded number of sessions and compares similarly to ProVerif for an unbounded number of sessions, with the advantage of a tighter treatment of bounded number of sessions. It can be applied to protocols with standard primitives and else branches.

7.2.2. Analysis of stateful security protocols

Participants: Vincent Cheval, Véronique Cortier, Jannik Dreier, Steve Kremer, Mathieu Turuani.

Many real-life protocols need to maintain a global state–such as counters, tables, or more generally, memory cells–that may be read and updated by parallel threads. Modelling such mutable, global state in protocols complicates the verification problem, in particular when analyzing an unbounded number of sessions.

The *SAPIC/TAMARIN* toolchain is one of the few tools that was designed to handle such global state. Dreier, Duménil (former intern in Pesto) and Kremer, in collaboration with Sasse (ETH Zurich, Switzerland) improve the underlying theory and the *TAMARIN* tool to allow for more general user-specified equational theories: the extension supports arbitrary convergent equational theories that have the finite variant property, making *TAMARIN* the first tool to support at the same time this large set of user-defined equational theories, protocols with global mutable state, an unbounded number of sessions, and complex security properties. The effectiveness of this generalization is demonstrated by analyzing several protocols that rely on blind signatures, trapdoor commitment schemes, and ciphertext prefixes that were previously out of scope. This work has been presented at POST'17 [27].

ProVerif is a very popular tool for the analysis of security protocols, that works very well in practice. However, in the case of protocols with global states, ProVerif typically fails in its analysis, due to its internal abstraction. Instead of designing a new ad-hoc procedure, we devise a generic transformation of the security properties queried to ProVerif. We prove the soundness of our transformation and implement it into a front-end GSVerif. Our experiments show that our front-end (combined with ProVerif) outperforms the few existing tools, both in terms of efficiency and protocol coverage. We successfully apply our tool to a dozen of protocols of the literature including a deployed voting and a payment protocol. This work is under submission.

7.2.3. Analysis of e-voting protocols

Participants: Véronique Cortier, Constantin-Catalin Dragan, Mathieu Turuani.

Cortier and Dragan provide the first machine-checked proof of privacy-related properties (including ballot privacy) for an electronic voting protocol in the computational model. They target the popular Helios family of voting protocols, for which they identify appropriate levels of abstractions to allow for simplification and convenient reuse of proof steps across many variations of the voting scheme. The resulting framework enables machine-checked security proofs for several hundred variants of Helios and should serve as a stepping stone for the analysis of further variations of the scheme. In addition, they highlight some of the lessons learned regarding the gap between pen-and-paper and machine-checked proofs, and report on the experience with formalizing the security of protocols at this scale. This work has been presented at S&P'17 [21].

Turuani and Cortier, in collaboration with Galindo (Univ Birmingham), have analysed the e-voting protocol developed by Scytl and planned to be deployed in Switzerland. The formal analysis of both privacy and individual verifiability has been conducted in ProVerif. It required to develop a crafty encoding of the security properties in order to avoid the limitations of ProVerif in the presence of global states (here, no revoting). This first encoding yielded the preliminary ideas for the GSVerif tool mentioned in the previous section. Such a formal analysis is required by the Swiss Chancellerie and has been accepted at EuroSP'18 [23].

Norway used e-voting in its last political election both in September 2011 and September 2013. The underlying protocol was also developed by Scytl. Cortier, in collaboration with Wiedling, has conducted a formal analysis (by hand) of vote privacy of this protocol, considering several corruption scenarios [13].

7.2.4. *Unification in Forward-Closed Theories*

Participant: Christophe Ringeissen.

In collaboration with Marshall (Univ Mary Washington, USA) and Erbatur (LMU, Germany), we investigate the unification problem in equational theories involving forward-closed convergent term rewrite systems. In the class of forward-closed theories, unification is decidable and finitary since a convergent term rewrite system has a finite forward-closure if and only if it has the finite variant property. Actually, forward-closed theories are syntactic theories admitting a terminating mutation-based unification procedure. This can be shown by reusing a mutation-based unification algorithm originally developed for equational theories saturated by paramodulation, since a forward-closed theory is indeed a sufficient condition to get soundness and completeness. Building on this fact we develop a new mutation-based unification algorithm which is simpler, with regard to conflicts and number of rules, than the first algorithm. We then use this simplified algorithm as a component to develop a new method that solves the unification problem in unions of forward-closed theories with non-disjoint theories. The resulting algorithm can be viewed as a terminating instance of a procedure initiated for hierarchical combination. This work has been presented at the workshop UNIF'17 [33].

7.2.5. *Analysis of Combinations of Protocols*

Participant: Jannik Dreier.

When trying to prove the security of a protocol, one usually analyzes the protocol in isolation, i.e., in a network with no other protocols. But in reality, there will be many protocols operating on the same network, maybe even sharing data including keys, and an intruder may use messages of one protocol to break another. We call that a multi-protocol attack. In this work, we tried to find such attacks using the *TAMARIN* prover. We analyzed both examples that were previously analyzed by hand or using other tools, and found novel attacks. This work was presented at FPS'17 [31].

7.3. Design

7.3.1. *E-voting protocols*

Participants: Véronique Cortier, Alicia Filipiak.

Building upon a recently proposed voting scheme, BeleniosRF, we design a new voting scheme that ensures both verifiability and privacy against a compromised voting machine, as well as a compromised voting server. It assumes that the voter has two devices: one computer for casting a vote and another device (typically a smartphone or a tablet) to, optionally, audit the material (a voting sheet) sent to the voter. Neither the computer nor the smartphone learns how the voter voted unless they collude. The resulting protocol has been formally analysed in ProVerif w.r.t. both verifiability and privacy. Analysing verifiability in ProVerif cannot be done directly as it would require counting. Instead, we propose a set of properties that can be handled by ProVerif and that entail verifiability. This work is one of the contribution of the thesis manuscript of Alicia Filipiak and will be submitted.

7.3.2. *Designing and proving an EMV-compliant payment protocol for mobile devices*

Participants: Véronique Cortier, Alicia Filipiak.

In collaboration with Gharout, Traoré and Florent (Orange Labs), we devised a payment protocol that can be securely used on mobile devices, even infected by malicious applications. Our protocol only requires a light use of Secure Elements, which significantly simplifies certification procedures and protocol maintenance. It is also fully compatible with the EMV-SDA protocol and allows off-line payments for the users. We provide a formal model and full security proofs of the protocol using the *TAMARIN* prover. This work has been presented at EuroS&P'17 [22].

7.3.3. Composition and design of PKIs

Participants: Vincent Cheval, Véronique Cortier.

In protocol analysis one makes the (strong) assumption that honestly generated keys are available to all parties and that the link between identities and public keys is fixed and known to everyone. The abstraction is grounded in solid intuition but there are currently no theoretical underpinnings to justify its use. Cheval and Cortier, in collaboration with Warinschi (Univ Bristol, UK), initiate a rigorous study of how to use PKIs within other protocols, securely. They first show that the abstraction outlined above is in general unsound by exhibiting a simple protocol which is secure with idealized key distribution but fails in the presence of more realistic PKI instantiation. Their main result is a generic composition theorem that identifies under which conditions protocols that require public keys can safely use any PKI protocol (which satisfies a security notion which we identify). Interestingly, unlike most existing composition results in symbolic models they do not require full tagging of the composed protocols. Furthermore, the results confirm the recommended practice that keys used in the PKI should not be used for any other cryptographic task. This work has been presented at CSF'17 [19].

7.3.4. Privacy Protection in Social Networks

Participants: Younes Abid, Hector Dang-Nhu, Andrii Dychka, Abdessamad Imine, Michaël Rusinowitch, Valentin Salquebre.

In order to demonstrate privacy threats in social networks we show how to infer user preferences by random walks in a multiple graph representing simultaneously attributes and relationships links. For the approach to scale in a first phase we reduce the space of attribute values by partition in balanced homogeneous clusters. Following the Deepwalk approach, the random walks are considered as sentences. Hence unsupervised learning techniques from natural languages processing can be employed in a second phase to deduce semantic similarities of some attributes. We conduct initial experiments on real datasets to evaluate our approach. This work was presented at DEXA'17 [15].

7.3.5. Compressed and Verifiable Filtering Rules in Software-defined Networking

Participants: Haftay Gebreslasie Abreha, Michaël Rusinowitch.

In a joint project with EPI Madynes and Cynapsys, we are starting to work on the design, implementation and evaluation of multi-masked techniques for building a compressed and a verifiable filtering rules in Software Defined Networks with the possibility of distributing the workload processing among several packet filtering devices operating in parallel.

PRIVATICS Project-Team

6. New Results

6.1. A refinement approach for the reuse of privacy risk analysis results

Participants: Daniel Le Métayer, Sourya joyee de.

With the adoption of the EU General Data Protection Regulation (GDPR), conducting a data protection impact assessment will become mandatory for certain categories of personal data processing. A large body of literature has been devoted to data protection impact assessment and privacy impact assessment. However, most of these papers focus on legal and organizational aspects and do not provide many details on the technical aspects of the impact assessment, which may be challenging and time consuming in practice. The general objective of [10] was to fill this gap and to propose a methodology which can be applied to conduct a privacy risk analysis in a systematic way, to use its results in the architecture selection process (following the privacy by design approach and to re-use its generic part for different products or deployment contexts. The proposed analysis proceeds in three broad phases: (1) a generic privacy risk analysis phase which depends only on the specifications of the system and yields generic harm trees; (2) an architecture-based privacy risk analysis which takes into account the definitions of the possible architectures of the system and refines the generic harm trees into architecture-specific harm trees. (3) a context-based privacy risk analysis which takes into account the context of deployment of the system (e.g., a casino, an office cafeteria, a school) and further refines the architecture-specific harm trees into context-specific harm trees. Context-specific harm trees can be used to take decisions about the most suitable architectures.

6.2. Interdisciplinarity in practice: Challenges and benefits for privacy research

Participant: Daniel Le Métayer.

The objective of this work was to draw the lessons learned from a project that involved security systems engineers, computer scientists, lawyers and social scientists. Since one of the goals of the project was to propose actual solutions following the privacy by design approach, its aim was to go beyond multidisciplinarity and build on the variety of expertise available in the consortium to follow a true interdisciplinary approach. We have described the challenges before analyzing the solutions adopted by the project to meet them and the outcomes and benefits of the approach. We have concluded the study with some lessons to be drawn from this experience and recommendations for future interdisciplinary projects.

6.3. Capacity: an abstract model of control over personal data

Participant: Daniel Le Métayer.

While the control of individuals over their personal data is increasingly seen as an essential component of their privacy, the word “control” is usually used in a very vague way, both by lawyers and by computer scientists. This lack of precision may lead to misunderstandings and makes it difficult to check compliance. To address this issue, we have proposed in [17] a formal framework based on capacities to specify the notion of control over personal data and to reason about control properties. We have illustrated our framework with social network systems and shown that it makes it possible to characterize the types of control over personal data that they provide to their users and to compare them in a rigorous way. This work will be presented at CODASPY 2018.

6.4. Privacy Risk Analysis to Enable Informed Privacy Settings

Participants: Daniel Le Métayer, Sourya joyee de.

We have proposed in [16] a method to enable better informed choices of privacy preferences or privacy settings by individuals. The method relies on a privacy risk analysis framework parameterized with privacy settings. The user can express his choices, visualize their impact on the privacy risks through a user-friendly interface, and decide to revise them as necessary to reduce risks to an acceptable level.

6.5. Secure electronic documents: is the centralisation of biometric data really inevitable? Inria Analysis Note

Participants: Claude Castelluccia, Daniel Le Métayer.

The decree of 28 October 2016 authorising the creation of a centralised file of "secure electronic documents" (TES) has raised a certain number of questions and concerns. The main aim put forward by the French government is the fight against identity fraud. However, the text of the decree also authorises certain accesses to the database by officers of the national police, national Gendarmerie and intelligence. Many voices have been raised to highlight the risks that such a centralised file could represent with regard to individual freedom, and particularly the invasion of citizens' privacy. The strengthening of the means to fight fraud (and, more generally, criminality) and the requirement to protect privacy are not necessarily in contradiction. However, in order to be able to reach a decision on the advantages and disadvantages of a management system for electronic documents, it seemed necessary to:

- Clearly define the desired functionalities and the advantages that can be expected from them, in particular with respect to the current situation and other solutions.
- Describe the technical solution chosen in a sufficiently precise way to enable its analysis.
- Rigorously analyse the risks of an invasion of privacy with regard to the expected benefits.

As a contribution to this debate, we have analyzed several architectures and alternative solutions which are described in an Inria Analysis Note [15].

6.6. Biometric Systems Private by Design: Reasoning about privacy properties of biometric system architectures

Participant: Daniel Le Métayer.

The goal of this was to show the applicability of the privacy by design approach to biometric systems and the benefit of using formal methods to this end. Starting from a general framework to define privacy architectures and to formally reason about their properties, we have described its adaptation to biometrics. The choice of particular techniques and the role of the components (central server, secure module, biometric terminal, smart card, etc.) in the architecture have a strong impact on the privacy guarantees provided by a biometric system. In the literature, some architectures have already been analysed in some way. However, the existing proposals were made on a case by case basis, which makes it difficult to compare them and to provide a rationale for the choice of specific options. In this work, we have described, on different architectures providing different levels of protection, how a general framework for the definition of privacy architectures can be used to specify the design options of a biometric systems and to reason about them in a formal way.

6.7. Wi-Fi and privacy

Participants: Mathieu Cunche, Célestin Matte.

As communications-enabled devices are becoming more and more ubiquitous, it becomes easier to track the movements of individuals through the radio signals broadcasted by their devices. While there is a strong interest for physical analytics platforms to leverage this information for many purposes, this tracking also threatens the privacy of individuals. To solve this issue, we propose a privacy-preserving solution for collecting aggregate mobility patterns while at the same time satisfying the strong guarantee of ε-differential privacy. More precisely, we introduce a sanitization mechanism for efficient, privacy-preserving and non-interactive approximate distinct counting for physical analytics based on perturbed Bloom filters. We also extend and generalize previous approaches for estimating distinct count of events and joint events (i.e., intersection, and more generally tout of -n cardinalities). Finally, we experimentally evaluate our approach and compare it to previous ones on a real dataset.

Wi-Fi signals emitted by mobile smartphones can be exploited to passively track users' mobility. Turning off the Wi-Fi interface of the device is often presented as a mean to evade those tracking systems. As a matter of fact this method is sometime suggested by the actors of the Wi-Fi tracking industry as a way to opt-out from those systems. The Android system features an option to enable or disable Wi-Fi on the device. However, disabling Wi-Fi through this option is not sufficient to prevent all Wi-Fi activity of the device. Based on measurements on a range of Android devices, we show in [18] that another option, called "Always allow scanning", when activated, makes a device send Wi-Fi frames which can be used to track this device, even if the Wi-Fi switch is off. This option is not clearly described in all Android versions, and sometimes even not deactivatable. Besides, the Google Maps application prompts the user to activate this option.

6.8. Towards Privacy-preserving Wi-Fi Analytics

Participants: Mathieu Cunche, Mohammad Alaggan.

A new technique enabling non-interactive (t, n)-incidence count estimation for indicator vectors ensuring Differential Privacy has been introduced. Given one or two differentially private indicator vectors, estimating the distinct count of elements in each and their intersection cardinality (equivalently, their inner product) have been studied in the literature, along with other extensions for estimating the cardinality set intersection in case the elements are hashed prior to insertion. The core contribution behind all these studies was to address the problem of estimating the Hamming weight (the number of bits set to one) of a bit vector from its differentially private version, and in the case of inner product and set intersection, estimating the number of positions which are jointly set to one in both bit vectors. We develop in [13] the most general case of estimating the number of positions which are set to one in exactly t out of n bit vectors (this quantity is denoted the (t, n)-incidence count), given access only to the differentially private version of those bit vectors. This means that if each bit vector belongs to a different owner, each can locally sanitize their bit vector prior to sharing it, hence the non-interactive nature of our algorithm. The newly introduced algorithm simultaneously estimates the (t, n)-incidence counts for all $t \in \{0, ..., n\}$. Upper and lower bounds to the estimation error have been derived. The lower bound is achieved by generalizing the limit of two-party differential privacy into n-party differential privacy, which is a contribution of independent interest. We prove that a lower bound on the additive error that must be incurred by any n-wise inner product of n mutually differentially-private bit vectors. Those results are very general and are not limited to differentially private bit vectors. They should apply to a large class of sanitization mechanism of bit vectors which depend on flipping the bits with a constant probability. Some potential applications for this technique include physical mobility analytics, call-detail-record analysis, and similarity metrics computation.

6.9. Towards Implicit Visual Memory-Based Authentication

Participant: Claude Castelluccia.

Selecting and remembering secure passwords puts a high cognitive burden on the user, which has adverse effects on usability and security. Authentication schemes based on implicit memory can relieve the user of the burden of actively remembering a secure password. In [8], we propose a new authentication scheme (MooneyAuth) that relies on implicitly remembering the content of previously seen Mooney images. These images are thresholded two-tone images derived from images containing single objects. Our scheme has two phases: In the enrollment phase, a user is presented with Mooney images, their corresponding original images, and labels. This creates an implicit link between the Mooney image and the object in the user's memory that serves as the authentication secret. In the authentication phase, the user has to label a set of Mooney images, a task that gets performed with substantially fewer mistakes if the images have been seen in the enrollment phase. We applied an information-theoretical approach to compute the eligibility of the user, based on which images were labeled correctly. This new dynamic scoring is substantially better than previously proposed static scoring by considering the surprisal of the observed events. We built a prototype and performed three experiments with 230 and 70 participants over the course of 264 and 21 days, respectively. We show that MooneyAuth outperforms current implicit memory-based schemes, and demonstrates a promising new approach for fallback authentication procedures on the Web.

6.10. MyAdChoices: Bringing transparency and control to online advertising

Participant: Claude Castelluccia.

The intrusiveness and the increasing invasiveness of online advertising have, in the last few years, raised serious concerns regarding user privacy and Web usability. As a reaction to these concerns, we have witnessed the emergence of a myriad of ad-blocking and antitracking tools, whose aim is to return control to users over advertising. The problem with these technologies, however, is that they are extremely limited and radical in their approach: users can only choose either to block or allow all ads. With around 200 million people regularly using these tools, the economic model of the Web—in which users get content free in return for allowing advertisers to show them ads—is at serious peril. In [3], we propose a smart Web technology that aims at bringing transparency to online advertising, so that users can make an informed and equitable decision regarding ad blocking. The proposed technology is implemented as a Web-browser extension and enables users to exert fine-grained control over advertising, thus providing them with certain guarantees in terms of privacy and browsing experience, while preserving the Internet economic model. Experimental results in a real environment demonstrate the suitability and feasibility of our approach, and provide preliminary findings on behavioral targeting from real user browsing profiles.

6.11. Differentially Private Mixture of Generative Neural Networks

Participant: Claude Castelluccia.

Generative models are used in a wide range of applications building on large amounts of contextually rich information. Due to possible privacy violations of the individuals whose data is used to train these models, however, publishing or sharing generative models is not always viable. In [4], we develop a novel technique for privately releasing generative models and entire high-dimensional datasets produced by these models. We model the generator distribution of the training data with a mixture of k generative neural networks. These are trained together and collectively learn the generator distribution of a dataset. Data is divided into k clusters, using a novel differentially private kernel k-means, then each cluster is given to separate generative neural networks, such as Restricted Boltzmann Machines or Variational Autoencoders, which are trained only on their own cluster using differentially private gradient descent. We evaluate our approach using the MNIST dataset, as well as call detail records and transit datasets, showing that it produces realistic synthetic samples, which can also be used to accurately compute arbitrary number of counting queries.

6.12. Revisiting Private Web Search using Intel SGX

Participant: Antoine Boutet.

The leakage of user search queries by search engines, which is at the heart of their economic model, makes private Web search an essential functionality to offer to those users that care about their privacy. Nowadays, there exists no satisfactory approach to enable users to access search engines in a privacy-preserving way. Existing solutions are either too costly due to the heavy use of cryptographic mechanisms (e.g., private information retrieval protocols), subject to attacks (e.g., Tor, TrackMeNot, GooPIR) or rely on weak adversarial models (e.g., PEAS). This work [6] introduces X-Search, a novel private Web search mechanism building on the disruptive software guard extensions (SGX) proposed by Intel. We compare X-Search to its closest competitors, Tor and PEAS using a dataset of real web search queries. Our evaluation shows that: (1) X-Search offers stronger privacy guarantees than its competitors as it operates under a stronger adversarial model; (2) it better resists state-of-the-art re-identification attacks; (3) from the performance perspective, X-Search outperforms its competitors both in terms of latency and throughput by orders of magnitude.

6.13. PULP: Achieving Privacy and Utility Trade-off in User Mobility Data

Participant: Antoine Boutet.

Leveraging location information in location-based services leads to improving service utility through geo-contextualization. However, this raises privacy concerns as new knowledge can be inferred from location records, such as user's home and work places, or personal habits. Although Location Privacy Protection Mechanisms (LPPMs) provide a means to tackle this problem, they often require manual configuration posing significant challenges to service providers and users. Moreover, their impact on data privacy and utility is seldom assessed. In [9], we present PULP, a model-driven system which automatically provides user-specific privacy protection and contributes to service utility via choosing adequate LPPM and configuring it. At the heart of PULP is nonlinear models that can capture the complex dependency of data privacy and utility for each individual user under given LPPM considered, i.e., Geo-Indistinguishability and Promesse. According to users' preferences on privacy and utility, PULP efficiently recommends suitable LPPM and corresponding configuration. We evaluate the accuracy of PULP's models and its effectiveness to achieve the privacy-utility trade-off per user, using four real-world mobility traces of 770 users in total. Our extensive experimentation shows that PULP ensures the contribution to location service while adhering to privacy constraints for a great percentage of users, and is orders of magnitude faster than non-model based alternatives.

6.14. The Pitfalls of Hashing for Privacy

Participants: Cédric Lauradoux, Mathieu Cunche, Levent Demir.

Boosted by recent legislations, data anonymization is fast becoming a norm. However, as of yet no generic solution has been found to safely release data. As a consequence, data custodians often resort to ad-hoc means to anonymize datasets. Both past and current practices indicate that hashing is often believed to be an effective way to anonymize data. Unfortunately, in practice it is only rarely effective. In [2], we expose the limits of cryptographic hash functions as an anonymization technique. Anonymity set is the best privacy model that can be achieved by hash functions. However, this model has several shortcomings. We provide three case studies to illustrate how hashing only yields a weakly anonymized data. The case studies include MAC and email address anonymization as well as the analysis of Google Safe Browsing.

6.15. Duck Attack on Accountable Distributed Systems

Participant: Cédric Lauradoux.

Accountability plays a key role in dependable distributed systems. It allows to detect, isolate and churn malicious/selfish nodes that deviate from a prescribed protocol. To achieve these properties, several accountable systems use at their core cryptographic primitives that produce non-repudiable evidence of inconsistent or incorrect behavior. In [11], we show how selfish and colluding nodes can exploit the use of cryptographic digests in accountability protocols to mount what we call a duck attack. In a duck attack, selfish and colluding nodes exploit the use of cryptographic digests to alter the transmission of messages while masquerading as honest entities. The end result is that their selfish behavior remains undetected. This undermines the security guarantees of the accountability protocols. We first discover the duck attack while analyzing PAG – a custom cryptographic protocol to build accountable systems presented at ICDCS 2016. We later discover that accountable distributed systems based on a secure log (essentially a hash-based data structure) are also vulnerable to the duck attack and apply it on AcTinG – a protocol presented at SRDS 2014. To defeat our attack, we modify the underlying secure log to have high-order dependency on the messages stored in it.

6.16. Less Latency and Better Protection with AL-FEC Sliding Window Codes: a Robust Multimedia CBR Broadcast Case Study

Participants: Vincent Roca, Belkacem Teibi.

Application-Level Forward Erasure Correction (AL-FEC) codes have become a key component of communication systems in order to recover from packet losses. This work analyzes the benefits of the AL-FEC codes based on a sliding encoding window (A.K.A. convolutional codes) for the reliable broadcast of real-time flows to a potentially large number of receivers over a constant bit rate channel. It first details the initialization of both sliding window codes and traditional block codes in order to keep the maximum AL-FEC decoding latency below a target latency budget. Then it presents detailed performance analyzes using official 3GPP mobility traces, representative of our use case which involves mobile receivers. This work highlights the major benefits of RLC codes, representative of sliding window codes, that outperform any block code, from Raptor codes (that are part of 3GPP MBMS standard) up to ideal MDS codes, both in terms of reduced added latency and improved robustness. It also demonstrates that our RLC codec features decoding speeds that are an order of magnitude higher than that of Raptor codes.

6.17. Coding for efficient Network Communications Research Group (NWCRG)

Participants: Vincent Roca, Belkacem Teibi.

In the context of the "Coding for efficient Network Communications IRTF Research Group (NWCRG) » (https://datatracker.ietf.org/rg/nwcrg/), several activities have been carried out. First of all, a recommended terminology for Network Coding concepts and constructs has been elaborated. It provides a comprehensive set of terms in order to avoid ambiguities in future Network Coding IRTF and IETF documents.

Then, in order to facilitate the use of Sliding Window Codes, such as RLC (see the above FEC Scheme) and RLNC codes (i.e., the well known codes for network coding applications, with potential re-encoding within the network), a work started that introduces a generic Application Programming Interface (API) for window-based FEC codes. This API is meant to be usable by any sliding window FEC code, independently of the FEC Scheme or network coding protocol that may rely on it. This API defines the core procedures and functions meant to control the codec (i.e., implementation of the FEC code), but leaves out all upper layer aspects (e.g., signalling) that are the responsibility of the application making use of the codec. A goal of this document is to pave the way for a future open-source implementation of such codes.

Finally, we started to work on the motivation and requirements for the use of Network Level Packet Erasure Coding to improve the performance of the QUIC protocol that is proposed a new transport protocol. The goal at this level is not specify a specific code but to list the salient features that a code should have in order to deal with know loss patterns on QUIC paths.

PROSECCO Project-Team

7. New Results

7.1. Verification of Security Protocols in the Symbolic Model

Participants: Bruno Blanchet, Marc Sylvestre.

The applied pi calculus is a widely used language for modeling security protocols, including as a theoretical basis of ProVerif. However, the seminal paper that describes this language [45] does not come with proofs, and detailed proofs for the results in this paper were never published. Martín Abadi, Bruno Blanchet, and Cédric Fournet wrote detailed proofs of all results of this paper. This work appears in the Journal of the ACM [12].

Marc Sylvestre improved the display of attacks in ProVerif, in particular by showing the computations performed by the attacker to obtain the messages sent in the attack, and by explaining why the found trace breaks the considered security property. He also developed an interactive simulator that allows the user to run the protocol step by step. The extended tool is available at http://proverif.inria.fr.

7.2. Symbolic and Computational Verification of Signal

Participants: Karthikeyan Bhargavan, Bruno Blanchet, Nadim Kobeissi.

We proposed a novel methodology that allows protocol designers, implementers, and security analysts to collaboratively verify a protocol using automated tools. The protocol is implemented in ProScript, a new domain-specific language that is designed for writing cryptographic protocol code that can both be executed within JavaScript programs and automatically translated to a readable model in the applied pi calculus. This model can then be analyzed symbolically using ProVerif to find attacks in a variety of threat models. The model can also be used as the basis of a computational proof using CryptoVerif, which reduces the security of the protocol to standard cryptographic assumptions. If ProVerif finds an attack, or if the CryptoVerif proof reveals a weakness, the protocol designer modifies the ProScript protocol code and regenerates the model to enable a new analysis. We demonstrated our methodology by implementing and analyzing two protocols: a variant of the popular Signal Protocol and TLS 1.3 Draft-18.

In our analysis of Signal, we used ProVerif and CryptoVerif to find new and previously-known weaknesses in the protocol and suggest practical countermeasures. Our ProScript protocol code is incorporated within the current release of Cryptocat, a desktop secure messenger application written in JavaScript. Our results indicate that, with disciplined programming and some verification expertise, the systematic analysis of complex cryptographic web applications is now becoming practical [33].

7.3. Symbolic and Computational Verification of TLS 1.3

Participants: Karthikeyan Bhargavan, Bruno Blanchet, Nadim Kobeissi.

We also applied our verification methodology to TLS 1.3, the next version of the Transport Layer Security (TLS) protocol. Its clean-slate design is a reaction both to the increasing demand for low-latency HTTPS connections and to a series of recent high-profile attacks on TLS. The hope is that a fresh protocol with modern cryptography will prevent legacy problems; the danger is that it will expose new kinds of attacks, or reintroduce old flaws that were fixed in previous versions of TLS. The protocol is nearing completion, and the working group has appealed to researchers to analyze the protocol before publication. We responded by presenting a comprehensive analysis of the TLS 1.3 Draft-18 protocol.

We seeked to answer three questions that had not been fully addressed in previous work on TLS 1.3: (1) Does TLS 1.3 prevent well-known attacks on TLS 1.2, such as Logjam or the Triple Handshake, even if it is run in parallel with TLS 1.2? (2) Can we mechanically verify the computational security of TLS 1.3 under standard (strong) assumptions on its cryptographic primitives? (3) How can we extend the guarantees of the TLS 1.3 protocol to the details of its implementations?

To answer these questions, we used our methodology for developing verified symbolic and computational models of TLS 1.3 hand-in-hand with a high-assurance reference implementation of the protocol. We presented symbolic ProVerif models for various intermediate versions of TLS 1.3 and evaluated them against a rich class of attacks to reconstruct both known and previously unpublished vulnerabilities that influenced the current design of the protocol. We presented a computational CryptoVerif model for TLS 1.3 Draft-18 and proved its security. We presented RefTLS, an interoperable implementation of TLS 1.0-1.3 in ProScript and automatically analyzed its protocol core by extracting a ProVerif model from its typed JavaScript code [24], [37]. This work was awarded the Distinguished Paper award at IEEE S&P 2017.

7.4. Verification of Avionic Security Protocols

Participant: Bruno Blanchet.

Within the ANR project AnaStaSec, we studied an air-ground avionic security protocol, the ARINC823 public key protocol [41]. We verified this protocol both in the symbolic model of cryptography, using ProVerif, and in the computational model, using CryptoVerif. While this study confirmed the main security properties of the protocol (entity and message authentication, secrecy), we found several weaknesses and imprecisions in the standard. We proposed fixes for these problems. This work appears in [27], [38].

We also verified the ATN Secure Dialogue protocol (ICAO 9880-IV [42]), which is currently under development. We verified it using ProVerif and CryptoVerif. While we confirmed the main security properties of the intended protocol, we found several incoherences, weaknesses, and imprecisions in the draft standard. We proposed fixes for these problems. We presented this work to the ICAO Secure Dialogue Subgroup (September 2017).

7.5. Design and Verification of next-generation protocols: identity, blockchains, and messaging

Participants: Harry Halpin, George Danezis [University College London], Carmela Troncoso [IMDEA].

We continued work on next-generation protocols via the NEXTLEAP project in 2017. The work started in 2016 to define the principles of design of decentralized protocols and a paper was published in the Privacy Enhancing Techologies Symposium as "Systematizing Decentralization and Privacy: Lessons from 15 years of research and deployments", which systematized over 180 papers from p2p to blockchains. We formally defined decentralization in terms of a distributed system operating in an adversarial environment, which we hope will be a foundational contribution to the field. NEXTLEAP also published a paper in ARES 2017 on how these principles can be applied to secure messaging systems, including the work of Prosecco on formalizing secure messaging as presented in EuroS&P 2017. NEXTLEAP had a successful launch event at Centre Pompidou, colocated with Eurocrypt, which was attended by a panel of prominent cryptographers (Phil Rogaway, Moti Yung, Tanja Lange, Daniel Bernstein) and members of the European Commission and European Parliament, attracting over 100 members of the general public to hear about Prosecco's research.

Building on the work on identity started in 2017, we finished the design of ClaimChain, the privacy-enhanced blockchain-based identity system, and work started on a F* implementation and scalability simulations. Unlike most blockchain systems that are public and are essentially replicated state machines, Claimchains use VRFs for privacy and do not require global consensus, instead allowing private linking between Claimchains and gossiping to maintain local consensus on secret material. We believe that this design may be the first workable approach to decentralizing PKI. Claimchains also use Merkle Trees for efficiency, and some of this library may end up as generally useful for F* programming after more development in 2018. Claimchain has yet to

be published in an academic venue, but it has already attracted considerable interest and was presented in the popular CCC security conference in Leipzig Germany. We also continued to raise the bar on security and privacy, hosting the first ever workshop on "Security and Privacy on the Blockchain" at EuroS&P 2017, which was sponsored by Blockstream. We expect the first formally verified blockchain system based on this design to be finished in 2018.

Another aspect of building next-generation protocols is to evaluate their usability. Prior studies have shown that users typically do not understand encryption and are even hostile to open-source code. However, these studies are typically done with students drawn for a general population, and in response Prosecco, in co-operation with sociologists from CNRS/Sorbonne, have started the largest-ever study of high-risk users from countries as diverse as Ukraine, Russia, Egypt and Tunisia. Preliminary results were presented at the European Usable Security (EuroUSEC) workshop, and already have attracted considerable attention from developers of secure messaging applications such as Signal and Briar. We hope that our findings on how users actually do group messaging and key verification will lead to changes in the underlying protocols.

Lastly, we continue to work with standards bodies in order to do security and privacy analysis of new protocols. For example, we have started formalizing W3C Web Authentication and inspecting its privacy properties, and our work on the lack of security in Semantic Web standards led to "Semantic Insecurity: Security and the Semantic Web" at ISWC 2017. Work on the security and privacy properties of the W3C Encrypted Media Extension led to an invited keynote at SPACE 2017.

Next year, we will finalize ClaimChain and add on the mix-network we have been developing over the last year, leading to a metadata-resistant and decentralized secure messaging application. We will work on spreading awareness of the importance of formally verified open standards as being necessary for the future of security, rather than closed-source solutions that may have backdoors and dangerous bugs that could cause severe economic damage if not fixed. To this end, we will work with ECRYPT CSA on the IACR Summer School of Societal and Business Impact of Cryptography, colocated with Real-World Crypto 2018, and co-organize an event at the European Commission and Parliament.

7.6. The F* programming language

Participants: Danel Ahman, Benjamin Beurdouche, Karthikeyan Bhargavan, Barry Bond [Microsoft Research], Tej Chajed [MIT], Antoine Delignat-Lavaud [Microsoft Research], Victor Dumitrescu, Cédric Fournet [Microsoft Research], Catalin Hritcu, Qunyan Mangus [Microsoft Research], Markulf Kohlweiss [Microsoft Research], Kenji Maillard, Asher Manning [McGill University], Guido Martínez [CIFASIS-CONICET Rosario], Zoe Paraskevopoulou [Princeton University], Clément Pit-Claudel [MIT], Jonathan Protzenko [Microsoft Research], Tahina Ramananandro [Microsoft Research], Aseem Rastogi [Microsoft Research], Jared Roesch [University of Washington], Nikhil Swamy [Microsoft Research], Christoph M. Wintersteiger [Microsoft Research], Santiago Zanella-Béguelin [Microsoft Research].

F* is an ML-like functional programming language aimed at program verification. Its type system includes polymorphism, dependent types, monadic effects, refinement types, and a weakest precondition calculus. Together, these features allow expressing precise and compact specifications for programs, including functional correctness and security properties. The F* type-checker aims to prove that programs meet their specifications using a combination of SMT solving and manual proofs. Programs written in F* can be translated to OCaml, F#, or C for execution.

The latest version of F* is written entirely in F*, and bootstraps in OCaml and F#. It is open source and under active development on http://github.com/FStarLang/FStar. A detailed description of this new F* version is available in a series of POPL papers [62], [22], [14].

The main ongoing use case of F* is building a verified, drop-in replacement for the whole HTTPS stack in Project Everest [25]. This includes verified implementations of TLS 1.2 and 1.3 including the underlying cryptographic primitives. Moreover, while F* is extracted to OCaml by default, we have devised a subset of F* that can be compiled to C for efficiency [18].

We released two versions of the software this year.

7.7. Micro-Policies

Participants: Arthur Azevedo de Amorim [University of Pennsylvania], Chris Casinghino [Draper Labs], André Dehon [University of Pennsylvania], Catalin Hritcu, Théo Laurent [ENS Paris], Benjamin Pierce [University of Pennsylvania], Howard Shrobe [MIT], Greg Sullivan [Dover Microsystems], Andrew Tolmach [Portland State University].

This year we obtained a new DARPA grant called SSITH/HOPE on "Advanced New Hardware Optimized for Policy Enforcement, A New HOPE". This grant is in the process of starting and our contribution will focus on devising a high-level micro-policy language and investigating micro-policies targetting today's most severe security vulnerabilities.

7.8. HACL*: A Verified Modern Cryptographic Library

Participants: Jean Karim Zinzindohoue, Karthikeyan Bhargavan, Jonathan Protzenko [Microsoft Research], Benjamin Beurdouche.

HACL* is a verified portable C cryptographic library that implements modern cryptographic primitives such as the ChaCha20 and Salsa20 encryption algorithms, Poly1305 and HMAC message authentication, SHA-256 and SHA-512 hash functions, the Curve25519 elliptic curve, and Ed25519 signatures.

HACL* is written in the F* programming language and then compiled to readable C code using the KreMLin tool [18]. The F* source code for each cryptographic primitive is verified for memory safety, mitigations against timing side-channels, and functional correctness with respect to a succinct high-level specification of the primitive derived from its published standard. The translation from F* to C preserves these properties and the generated C code can itself be compiled via the CompCert verified C compiler or mainstream compilers like GCC or CLANG. When compiled with GCC on 64-bit platforms, our primitives are as fast as the fastest pure C implementations in OpenSSL and Libsodium, significantly faster than the reference C code in TweetNaCl, and between 1.1x-5.7x slower than the fastest hand-optimized vectorized assembly code in the SUPERCOP benchmark test-suite.

HACL* implements the NaCl cryptographic API and can be used as a drop-in replacement for NaCl libraries like Libsodium and TweetNaCl. HACL* provides the cryptographic components for a new mandatory ciphersuite in TLS 1.3 and is being developed as the main cryptographic provider for the miTLS verified implementation. Primitives from HACL* have now been integrated within Mozilla's NSS cryptographic library. Our results show that writing fast, verified, and usable C cryptographic libraries is now practical.

This work appeared at the ACM CCS conference [36] and all our software is publicly available and in active development on GitHub.

7.9. miTLS: A Verified TLS Implementation

Participants: Karthikeyan Bhargavan, Antoine Delignat-Lavaud [Microsoft Research], Cédric Fournet [Microsoft Research], Markulf Kohlweiss [Microsoft Research], Jianyang Pan, Jonathan Protzenko [Microsoft Research], Aseem Rastogi [Microsoft Research], Nikhil Swamy [Microsoft Research], Santiago Zanella-Béguelin [Microsoft Research], Jean Karim Zinzindohoue.

The record layer is the main bridge between TLS applications and internal sub-protocols. Its core functionality is an elaborate authenticated encryption: streams of messages for each sub-protocol (handshake, alert, and application data) are fragmented, multiplexed, and encrypted with optional padding to hide their lengths. Conversely, the sub-protocols may provide fresh keys or signal stream termination to the record layer.

Compared to prior versions, TLS 1.3 discards obsolete schemes in favor of a common construction for Authenticated Encryption with Associated Data (AEAD), instantiated with algorithms such as AES-GCM and ChaCha20-Poly1305. It differs from TLS 1.2 in its use of padding, associated data and nonces. It encrypts the content-type used to multiplex between sub-protocols. New protocol features such as early application data (0-RTT and 0.5-RTT) and late handshake messages require additional keys and a more general model of stateful encryption.

As part of the miTLS project, we built and verified a reference implementation of the TLS record layer and its cryptographic algorithms in F*. We reduced the high-level security of the record layer to cryptographic assumptions on its ciphers. Each step in the reduction is verified by typing an F* module; when the step incurs a security loss, this module precisely captures the corresponding game-based security assumption.

We computed concrete security bounds for the AES-GCM and ChaCha20-Poly1305 ciphersuites, and derived recommended limits on sent data before re-keying. Combining our functional correctness and security results, we obtained the first verified implementation of the main TLS 1.3 record ciphers. We plugged our implementation into an existing TLS library and confirmed that the combination interoperates with Chrome and Firefox, and thus that experimentally the new TLS record layer (as described in RFCs and cryptographic standards) is provably secure.

This work appeared at IEEE S&P 2017 [26] and our verified software is publicly available and actively developed on GitHub.

7.10. A Cryptographic Analysis of Content Delivery of TLS

Participants: Karthikeyan Bhargavan, Ioana Boureanu [University of Surrey], Pierre-Alain Fouque [University of Rennes 1/IRISA], Cristina Onete [University of Rennes 1/IRISA], Benjamin Richard [Orange Labs Chatillon].

The Transport Layer Security (TLS) protocol is designed to allow two parties, a client and a server, to communicate securely over an insecure network. However, when TLS connections are proxied through an intermediate middlebox, like a Content Delivery Network (CDN), the standard end-to-end security guarantees of the protocol no longer apply.

As part of the SafeTLS project, we investigated the security guarantees provided by Keyless SSL, a CDN architecture currently deployed by CloudFlare that composes two TLS 1.2 handshakes to obtain a proxied TLS connection. We demonstrated new attacks that show that Keyless SSL does not meet its intended security goals. We argued that proxied TLS handshakes require a new, stronger, 3-party security definition, and we presented one.

We modified Keyless SSL and proved that our modifications guarantee this notion of security. Notably, we showed that secure proxying in TLS 1.3 is computationally lighter and requires simpler assumptions on the certificate infrastructure than our proposed fix for Keyless SSL. Our results indicate that proxied TLS architectures, as currently used by a number of CDNs, may be vulnerable to subtle attacks and deserve close attention [39].

TAMIS Team

7. New Results

7.1. Results for Axis 1: Vulnerability analysis

7.1.1. Statistical Model Checking of LLVM Code

Participants: Axel Legay, Louis-Marie Traonouez.

We have extended PLASMA Lab statistical model-checker with a new plugin that allows to simulate LLVM bitcode. The plugin is based on an external simulator LODIN. This simulator implements a probabilistic semantics for a LLVM program. At its core the semantics consist of the LLVM program given as a labelled transition system. The labels are function calls to an environment that implements functions outside the LLVM core language. The environment is also responsible for assigning probabilities to individual transitions

By interfacing the LODIN simulator with PLASMA Lab we can apply all the statistical model-checking algorithms provided by PLASMA Lab, including rare events verification algorithms like importance splitting. We have applied LODIN and PLASMA Lab to several case studies, including the analysis of some security vulnerability, like the PTrace privilege escalation attack that could be performed on earlier versions of the Linux Kernel. This work has been submitted to a conference this year [61], and is currently under review.

[61] We present our work in providing Statistical Model Checking for programs in LLVM bitcode. As part of this work we develop a semantics for programs that separates the program itself from its environment. The program interact with the environment through function calls. The environment is furthermore allowed to perform actions that alter the state of the C-program-useful for mimicking an interrupt system. On top of this semantics we build a probabilistic semantics and present an algorithm for simulating traces under that semantics. This paper also includes the development of the new tool component Lodin that provides a statistical model checking infrastructure for LLVM programs. The tool currently implement standard Monte Carlo algorithms and a simulator component to manually inspect the behaviour of programs. The simulator also proves useful in one of our other main contributions; namely producing the first tool capable of doing importance splitting on LLVM code. Importance splitting is implemented by integrating Lodin with the existing statistical model checking tool Plasma-Lab.

7.1.2. Verification of IKEv2 protocol

Participants: Axel Legay, Tristan Ninet, Louis-Marie Traonouez, Olivier Zendra.

The IKEv2 (Internet Key Exchange version 2) protocol is the authenticated key-exchange protocol used to set up secure communications in an IPsec (Internet Protocol security) architecture. It guarantees security properties like mutual-authentication and secrecy of the exchanged key. To obtain an IKEv2 implementation as secure as possible, we use model checking to verify the properties on the protocol specification, and smart fuzzing to test the implementation, and try to detect implementation flaws like buffer overflows or memory leaks.

Two weaknesses had previously been found in the specification, but were harmless. We showed that the first weakness does not actually exist. We demonstrated that the second weakness is not harmless, and we designed a Denial-of-Service attack that exploits it, the deviation attack. As a counter-measure, we propose a modification of IKEv2, and use model checking to prove that the modified version is secure.

This work is being prepared for responsive disclosure and publication.

7.1.3. High-Level Frameworks for Scheduling Systems

Participants: Mounir Chadli, Axel Legay, Louis-Marie Traonouez.

Formal model-based techniques are more and more used for the specification and verification of scheduling systems. These techniques allow to consider complex scheduling policies beyond the scope of classical analytical techniques. For instance, hierarchical scheduling systems (HSS) integrates a number of components into a single system running on one execution platform. Hierarchical scheduling systems have been gaining more attention by automotive and aircraft manufacturers because they are practical in minimizing the cost and energy of operating applications. Model-based techniques can also be used to solve new problems like energy optimization or runtime monitoring. However, one limitation of formal model-based approaches is that they require high technical knowledge about the formalims and tools used to design models and write properties.

In a previous work [62], we have presented a model-based framework for the verification of HSS. It is based on a stochastic extension of timed automata and statistical model checking with the tool UPPAAL. We have also developed a graphical high-level language to represent complex hierarchical scheduling systems. To bridge the gap between the formalisms, we exploit Cinco, a generator for domain specific modeling tools to generate an interface between this language and the one of UPPAAL. Cinco allows to specify the features of a graphical interface in a compact meta-model language. This is a flexible approach that could be extended to any formal model of scheduling problem.

We have extended the previous work in journal paper [55] published this year, where we provide another high-level framework for the verification of energy-aware scheduling systems. We also present two new analysis techniques. One that performs runtime monitoring in order to detect alarming change in the scheduling system, and one that performs energy optimization.

[55] Over the years, schedulability of Cyber-Physical Systems (CPS) has mainly been performed by analytical methods. These techniques are known to be effective but limited to a few classes of scheduling policies. In a series of recent work, we have shown that schedulability analysis of CPS could be performed with a model-based approach and extensions of verification tools such as UPPAAL. One of our main contributions has been to show that such models are flexible enough to embed various types of scheduling policies, which goes beyond those in the scope of analytical tools.

However, the specification of scheduling problems with model-based approaches requires a substantial modeling effort, and a deep understanding of the techniques employed in order to understand their results. In this paper we propose simplicity-driven high-level specification and verification frameworks for various scheduling problems. These frameworks consist of graphical and user-friendly languages for describing scheduling problems. The high-level specifications are then automatically translated to formal models, and results are transformed back into the comprehensible model view. To construct these frameworks we exploit a meta-modeling approach based on the tool generator Cinco.

Additionally we propose in this paper two new techniques for scheduling analysis. The first performs runtime monitoring using the CUSUM algorithm to detect alarming change in the system. The second performs optimization using efficient statistical techniques. We illustrate our frameworks and techniques on two case studies.

7.1.4. Side-channel Analysis of Cryptographic Substitution Boxes

Participants: Axel Legay, Annelie Heuser.

With the advent of the Internet of Things, we are surrounded with smart objects (aka things) that have the ability to communicate with each other and with centralized resources. The two most common and widely noticed artefacts are RFID and Wireless Sensor Networks which are used in supply-chain management, logistics, home automation, surveillance, traffic control, medical monitoring, and many more. Most of these applications have the need for cryptographic secure components which inspired research on cryptographic algorithms for constrained devices. Accordingly, lightweight cryptography has been an active research area over the last 10 years. A number of innovative ciphers have been proposed in order to optimize various performance criteria and have been subject to many comparisons. Lately, the resistance against side-channel attacks has been considered as an additional decision factor.

Side-channel attacks analyze physical leakage that is unintentionally emitted during cryptographic operations in a device (e.g., power consumption, electromagnetic emanation). This side-channel leakage is statistically dependent on intermediate processed values involving the secret key, which makes it possible to retrieve the secret from the measured data.

Side-channel analysis (SCA) for lightweight ciphers is of particular interest not only because of the apparent lack of research so far, but also because of the interesting properties of substitution boxes (S-boxes). Since the nonlinearity property for S-boxes usually used in lightweight ciphers (i.e., 4×4) can be maximally equal to 4, the difference between the input and the output of an S-box is much smaller than for instance for AES. Therefore, one could conclude that from that aspect, SCA for lightweight ciphers must be more difficult. However, the number of possible classes (e.g., Hamming weight (HW) or key classes) is significantly lower, which may indicate that SCA must be easier than for standard ciphers. Besides the difference in the number of classes and consequently probabilities of correct classification, there is also a huge time and space complexity advantage (for the attacker) when dealing with lightweight ciphers.

In [65], [64] we give a detailed study of lightweight ciphers in terms of side-channel resistance, in particular for software implementations. As a point of exploitation we concentrate on the non-linear operation (S-box) during the first round. Our comparison includes SPN ciphers with 4-bit S-boxes such as KLEIN, PRESENT, PRIDE, RECTANGLE, Mysterion as well as ciphers with 8-bit S-boxes: AES, Zorro, Robin. Furthermore, using simulated data for various signal-to-noise ratios (SNR) we present empirical results for Correlation Power Analysis (CPA) and discuss the difference between attacking 4-bit and 8-bit S-boxes.

An extension of this work is given in [10]. We investigate whether side-channel analysis is easier for lightweight ciphers than e.g. for AES. We cover both profiled and non-profiled techniques where we are interested in recovering secret (round)keys or intermediate states. In the case of non-profiled attacks, we evaluate a number of S-boxes appearing in lightweight ciphers using the confusion coefficient and empirical simulations.

First, we investigate in the scenario where the attacker targets the first round and thus exploits the S-box computation. We observe that the 8-bit S-boxes from AES, Zorro, and Robin perform similarly, whereas for 4-bit S-boxes we have a clear ranking, with the S-box of Piccolo being the weakest to attack and the S-box of KLEIN and Midori (1) the hardest. Interestingly, when considering the last round and thus the inverse S-box operation the ranking changes such that Mysterion is the weakest and PRESENT/LED is the most side-channel resistant cipher from the ones investigated. Moreover, we could observe that attacking the last round is equal or less efficient for all considered ciphers. Finally, we use the information gained from both rounds together, where this approach is of interest when the cipher does not use round keys from a key scheduling algorithm but rather uses the same (or a straightforward computable) key in each round. LED fulfils this requirement. For a reasonable low SNR, to reach a success rate of 0.9 an attack on both rounds only requires 100 traces, whereas an attack using the first round requires 200 traces and on the last 400 traces. This example highlights the important role the confusion coefficient (relationship between predicted intermediate states under a leakage model from different key hypotheses), and that not only the SNR (even if low) is a key factor influencing the success rate. Additionally, our result show that we cannot conclude that the 4-bit S-boxes are generally significantly less resistant than the investigated 8-bit S-boxes. In particular, when considering inverse S-boxes we showed that 4-bit S-boxes may be more resistant.

For profiled attacks, we analyze several machine learning techniques to recover 4-bit and 8-bit intermediate states. Our results show that attacking 4-bit is somewhat easier than attacking 8-bit, with the difference mainly stemming from the varying number of classes in one or the other scenario. Still, that difference is not so apparent as one could imagine. Since we work with only a single feature and yet obtain a good accuracy in a number of test scenarios, we are confident (as our experiments also confirm) that adding more features will render classification algorithms even more powerful, which will result in an even higher accuracy. Finally, we did not consider any countermeasures for the considered lightweight algorithms, since the capacity for adding countermeasures is highly dependent on the environment (which we assume to be much more constrained than in the case of AES). However, our results show that a smart selection of S-boxes results in an inherent resilience (especially for 4-bit S-boxes). Moreover, we show that in case of highly restricted devices, in which

countermeasures on the whole cipher are not practically feasible, a designer may choose to only protect the weakest round (first round) in the cipher to increase the side-channel resistant until a certain limit.

Our work in [23] concentrates on how to improve SCA resilience of ciphers without imposing any extra cost. This is possible by considering the inherent resilience of ciphers. We particularly concentrate on block ciphers which utilize S-boxes and therefore study the resilience of S-boxes against side-channel attacks. When discussing how to improve side-channel resilience of a cipher, an obvious direction is to use various masking or hiding countermeasures. However, such schemes come with a cost, e.g. an increase in the area and/or reduction of the speed. When considering lightweight cryptography and various constrained environments, the situation becomes even more difficult due to numerous implementation restrictions. However, some options are possible like using S-boxes that are easier to mask or (more on a fundamental level), using S-boxes that possess higher inherent side-channel resilience. In [23] we investigate what properties should an S-box possess in order to be more resilient against side-channel attacks. Moreover, we find certain connections between those properties and cryptographic properties like nonlinearity and differential uniformity. Finally, to strengthen our theoretical findings, we give an extensive experimental validation of our results.

[64] Side-channel Analysis of Lightweight Ciphers: Current Status and Future Directions
[65] Side-channel Analysis of Lightweight Ciphers: Does Lightweight Equal Easy?
[10] Lightweight Ciphers and their Side-channel Resilience.
[23] Trade-Offs for S-Boxes: Cryptographic Properties and Side-Channel Resilience
[24] Do we need a holistic approach for the design of secure IoT systems? hal-01628683

7.1.5. New Advances on Side-channel Distinguishers

Participants: Axel Legay, Annelie Heuser.

[16] *Template Attack vs Bayes Classifier*

Side-channel attacks represent one of the most powerful category of attacks on cryptographic devices with profiled attacks in a prominent place as the most powerful among them. Indeed, for instance, template attack is a well-known real-world attack that is also the most powerful attack from the information theoretic perspective. On the other hand, machine learning techniques have proven their quality in a numerous applications where one is definitely side-channel analysis. As one could expect, most of the research concerning supervised machine learn- ing and side-channel analysis concentrated on more powerful machine learning techniques. Although valid from the practical perspective, such attacks often remain lacking from the more theoretical side. In this paper, we investigate several Bayes classifiers, which present simple supervised techniques that have significant similarities with the template attack. More specifically, our analysis aims to investigate what is the influence of the feature (in)dependence in datasets with different amount of noise and to offer further insight into the efficiency of machine learning for side-channel analysis.

[46] *Side-channel analysis and machine learning: A practical perspective* The field of side-channel analysis has made significant progress over time. Analyses are now used in practice in design companies as well as in test laboratories, and the security of products against side-channel attacks has significantly improved. However, there are still some remaining issues to be solved for analyses to be more effective. Side-channel analysis ac- tually consists of two steps, commonly referred to as identification and exploitation. The identification consists of understanding the leakage in order to set up a relevant attack. On the other hand, the exploitation consists of using the identified leakages to extract the secret key. In scenarios where the model is poorly known, it can be approximated in a profiling phase. There, machine learning techniques are gaining value. In this paper, we conduct extensive analysis of several machine learning techniques, showing the importance of proper parameter tuning and training. In contrast to what is perceived as common knowledge in unrestricted scenarios, we show that some machine learning techniques can significantly outperform template attack when properly used. We therefore stress that the traditional worst case security assessment of cryptographic implementations that includes mainly template attacks might not be accurate enough. Besides that, we present a new measure called the Data Confusion Factor that can be used to assess how well machine learning techniques will perform on a certain dataset.

[30] *Codes for Side-Channel Attacks and Protections*

This article revisits side-channel analysis from the standpoint of coding theory. On the one hand, the attacker is shown to apply an optimal decoding algorithm in order to recover the secret key from the analysis of the side-channel. On the other hand, the side-channel protections are presented as a coding problem where the information is mixed with randomness to weaken as much as possible the sensitive information leaked into the side-channel. Therefore, the field of side-channel analysis is viewed as a struggle between a coder and a decoder. In this paper, we focus on the main results obtained through this analysis. In terms of attacks, we discuss optimal strategy in various practical contexts, such as type of noise, dimensionality of the leakage and of the model, etc. Regarding countermeasures, we give a formal analysis of some masking schemes.

[38] *Climbing Down the Hierarchy: Hierarchical Classification for Machine Learning Side-Channel Attacks*

Machine learning techniques represent a powerful paradigm in side-channel analysis, but they come with a price. Selecting the appropriate algorithm as well as the parameters can sometimes be a difficult task. Nevertheless, the results obtained usually justify such an effort. However, a large part of those results use simplification of the data relation and in fact do not consider allthe available information. In this paper, we analyze the hierarchical relation between the data and propose a novel hierarchical classification approach for side-channel analysis. With this technique, we are able to introduce two new attacks for machine learning side-channel analysis: Hierarchical attack and Structured attack. Our results show that both attacks can outperform machine learning techniques using the traditional approach as well as the template attack regarding accuracy. To support our claims, we give extensive experimental results and discuss the necessary conditions to conduct such attacks.

[14] *Stochastic Collision Attack*

On the one hand, collision attacks have been introduced in the context of side-channel analysis for attackers who exploit repeated code with the same data without having any knowledge of the leakage model. On the other hand, stochastic attacks have been introduced to recover leakage models of internally processed intermediate secret variables. Both techniques have shown advantages and intrinsic limitations. Most collision attacks, for instance, fail in exploiting all the leakages (e.g., only a subset of matching samples are analyzed), whereas stochastic attacks cannot involve linear regression with the full basis (while the latter basis is the most informative one). In this paper, we present an innovative attacking approach, which combines the flavors of stochastic and collision attacks. Importantly, our attack is derived from the optimal distinguisher, which maximizes the success rate when the model is known. Notably, we develop an original closed-form expression, which shows many benefits by using the full algebraic description of the leakage model. Using simulated data, we show in the unprotected case that, for low noise, the stochastic collision attack is superior to the state of the art, whereas asymptotically and thus, for higher noise, it becomes equivalent to the correlation-enhanced collision attack. Our so-called stochastic collision attack is extended to the scenario where the implementation is protected by masking. In this case, our new stochastic collision attack is more efficient in all scenarios and, remarkably, tends to the optimal distinguisher. We confirm the practicability of the stochastic collision attack thanks to experiments against a public data set (DPA contest v4). Furthermore, we derive the stochastic collision attack in case of zero-offset leakage that occurs in protected hardware implementations and use simulated data for comparison. Eventually, we underline the capability of the new distinguisher to improve its efficiency when the attack multiplicity increases.

[15] *Optimal side-channel attacks for multivariate leakages and multiple models*

Side-channel attacks allow to extract secret keys from embedded systems like smartcards or smartphones. In practice, the side-channel signal is measured as a trace consisting of several samples. Also, several sensitive bits are manipulated in parallel, each leaking differently. Therefore, the informed attacker needs to devise side-channel distinguishers that can handle both multivariate leak-

ages and multiple models. In the state of the art, these two issues have two independent solutions: on the one hand, dimensionality reduction can cope with multivariate leakage; on the other hand, online stochastic approach can cope with multiple models. In this paper, we combine both solutions to derive closed-form expressions of the resulting optimal distinguisher in terms of matrix operations, in all situations where the model can be either profiled offline or regressed online. Optimality here means that the success rate is maximized for a given number of traces. We recover known results for uni- and bivariate models (including correlation power analysis) and investigate novel distinguishers for multiple models with more than two parameters. In addition, following ideas from the AsiaCrypt?2013 paper ?Behind the Scene of Side-Channel Attacks,? we provide fast computation algorithms in which the traces are accumulated prior to computing the distinguisher values.

[39] *Stochastic Side-Channel Leakage Analysis via Orthonormal Decomposition*

Side-channel attacks of maximal efficiency require an accurate knowledge of the leakage function. Template attacks have been introduced by Chari et al. at CHES 2002 to estimate the leakage function using available training data. Schindler et al. noticed at CHES 2005 that the complexity of profiling could be alleviated if the evaluator has some prior knowledge on the leakage function. The initial idea of Schindler is that an engineer can model the leakage from the structure of the circuit. However, for some thin CMOS technologies or some advanced countermeasures, the engineer intuition might not be sufficient. Therefore, inferring the leakage function based on profiling is still important. In the state-of-the-art, though, the profiling stage is conducted based on a linear regression in a non-orthonormal basis. This does not allow for an easy interpretation because the components are not independent. In this paper, we present a method to characterize the leakage based on a Walsh-Hadamard orthonormal basis with staggered degrees, which allows for direct interpretations in terms of bits interactions. A straightforward application is the characterization of a class of devices in order to understand their leakage structure. Such information is precious for designers and also for evaluators, who can devise attack bases relevantly.

[17] *On the optimality and practicability of mutual information analysis in some scenarios*

The best possible side-channel attack maximizes the success rate and would correspond to a maximum likelihood (ML) distinguisher if the leakage probabilities were totally known or accurately estimated in a profiling phase. When profiling is unavailable, however, it is not clear whether Mutual Information Analysis (MIA), Correlation Power Analysis (CPA), or Linear Regression Analysis (LRA) would be the most successful in a given scenario. In this paper, we show that MIA coincides with the maximum likelihood expression when leakage probabilities are replaced by online estimated probabilities. Moreover, we show that the calculation of MIA is lighter that the computation of the maximum likelihood. We then exhibit two case-studies where MIA outperforms CPA. One case is when the leakage model is known but the noise is not Gaussian. The second case is when the leakage model is partially unknown and the noise is Gaussian. In the latter scenario MIA is more efficient than LRA of any order.

[59] *On the Relevance of Feature Selection for Profiled Side-channel Attacks*

In the process of profiled side-channel analysis there is a number of steps one needs to make. One important step that is often conducted without a proper attention is selection of the points of interest (features) within the side-channel measurement trace. Most of the related work start with an assumption that the features are selected and various attacks are then considered and compared to find the best approach. In this paper, we concentrate on the feature selection step and show that if a proper selection is done, most of the attack techniques offer satisfactory results. We investigate how more advanced feature selection techniques stemming from the machine learning domain can be used to improve the side-channel attack efficiency. Our results show that the so-called Hybrid feature selection methods result in the best classification accuracy over a wide range of test scenarios and number of features selected.

[60] *Profiled SCA with a New Twist: Semi-supervised Learning*

Profiled side-channel attacks represent the most powerful category of side-channel attacks. In this context, the attacker gains ac- cess of a profiling device to build a precise model which is used to attack another device in the attacking phase. Mostly, it is assumed that the attacker has unlimited capabilities in the profiling phase, whereas the attacking phase is very restricted. We step away from this assumption and consider an attacker who is restricted in the profiling phase, while the attacking phase is less limited as in the traditional view. Clearly, in general, the attacker is not hindered to exchange any available knowledge between the profiling and attacking phase. Accordingly, we propose the concept of semi-supervised learning to side-channel analysis, in which the attacker uses the small amount of labeled measurements from the profiling phase as well as the unlabeled measurements from the attacking phase to build a more reliable model. Our results show that semi-supervised learning is beneficial in many scenarios and of particular interest when using template attack and its pooled version as side-channel attack techniques. Besides stating our results in varying scenarios, we discuss more general conclusions on semi-supervised learning for SCA that should help to transfer our observations to other settings in SCA.

7.1.6. Side-channel analysis on post-quantum cryptography

Participants: Axel Legay, Annelie Heuser, Tania Richmond, Martin Moreau.

In recent years, there has been a substantial amount of research on quantum computers ? machines that exploit quantum mechanical phenomena to solve mathematical problems that are difficult or intractable for conventional computers. If large-scale quantum computers are ever built, they will be able to break many of the public-key cryptosystems currently in use. This would seriously compromise the confidentiality and integrity of digital communications on the Internet and elsewhere. The goal of post-quantum cryptography (also called quantum-resistant cryptography) is to develop cryptographic systems that are secure against both quantum and classical computers, and can interoperate with existing communications protocols and networks. At present, there are several post-quantum cryptosystems that have been proposed: lattice-based, code-based, multivariate cryptosystems, hash-based signatures, and others. However, for most of these proposals, further research is needed in order to gain more confidence in their security and to improve their performance. Our interest lies in particular on the side-channel analysis and resistance of these post-quantum schemes. We first focus on code-based cryptography and then extend our analysis to find common vulnerabilities between different families of post-quantum crypto systems.

7.1.7. Binary Code Analysis: Formal Methods for Fault Injection Vulnerability Detection

Participants: Axel Legay, Thomas Given-Wilson, Annelie Heuser, Nisrine Jafri, Jean-Louis Lanet.

Formal methods such as model checking provide a powerful tool for checking the behaviour of a system. By checking the properties that define correct system behaviour, a system can be determined to be correct (or not).

Increasingly fault injection is being used as both a method to attack a system by a malicious attacker, and to evaluate the dependability of the system. By finding fault injection vulnerabilities in a system, the resistance to attacks or faults can be detected and subsequently addressed.

A process is presented that allows for the automated simulation of fault injections. This process proceeds by taking the executable binary for the system to be tested, and validating the properties that represent correct system behaviour using model checking. A fault is then injected into the executable binary to produce a mutant binary, and the mutant binary is model checked also. A different result to the validation of the executable binary in the checking of the mutant binary indicates a fault injection vulnerability.

This process has been automated with existing tools, allowing for easy checking of many different fault injection attacks and detection of fault injection vulnerabilities. This allows for the detection of fault injection vulnerabilities to be fully automated, and broad coverage of the system to be formally shown.

The work is implemented in the SimFi tool.

[56] (J; submitted) Fault injection has increasingly been used both to attack software applications, and to test system robustness. Detecting fault injection vulnerabilities has been approached with a variety of different but limited methods. This paper proposes an extension of a recently published general model checking based process to detect fault injection vulnerabilities in binaries. This new extension makes the general process scalable to real-world implementations which is demonstrated by detecting vulnerabilities in different cryptographic implementations.

7.1.8. Security at the hardware and software boundaries

Participants: Axel Legay, Jean-Louis Lanet, Ronan Lashermes, Kevin Bukasa, Hélène Le Bouder.

7.1.8.1. Side-channel attacks (SCA)

SCA exploit the reification of a computation through its physical dimensions (current consumption, EM emission, etc.). Focusing on Electromagnetic Analyses (EMA), such analyses have mostly been considered on low-end devices: smartcards and micro-controllers. In the wake of recent works, we analyze the effects of a modern micro architecture [31] on the efficiency of EMA (here Correlation Power Analysis and template attacks). We show that despite the difficulty to synchronize the measurements, the speed of the targeted core and the activity of other cores on the same chip can still be accommodated. Finally, we confirm that enabling the secure mode of TrustZone (a hardware-assisted software countermeasure) has no effect whatsoever on the EMA efficiency. Therefore, critical applications in TrustZone are not more secure than in the normal world with respect to EMA, in accordance with the fact that it is not a countermeasure against physical attacks. We hint that such techniques may be more common in the future to overcome the true difficulty with high-end devices: dealing with time precision (problem even worse with an OS or a virtual machine). Here again TrustZone or the activity of other cores have no incidence. But with these attacks, managing the big amount of data generated by our measures may prove to be the limiting factor, requiring better computing resources.

We investigate the way the compiler works and new attack paths have been discovered. In particular we demonstrated experimentally on an ARM7m the possibility to execute arbitrary code, generate buffer overflow even in presence of compiler assisted canary and ROP attacks. This raises a new challenge: any code fragment of an embedded program is sensitive to a fault attack. Thus an attacker increases the success rate of its attack while targeting a non sensitive part of the program for the injection. Then it becomes easy to extract security materials from the device. Then, the verification of the absence of a potential vulnerability must be checked on the whole program and not only on the cryptographic primitives. Thus the prevention analysis that was possible thanks to formal methods becomes unreachable with these new attack paths [40].

7.1.8.2. SCA based fuzzer

One of the main challenges during the development of system is to give a proof of evidence that its functionalities are correctly implemented and that no vulnerability remains. This objective is mostly achieved via testing techniques, which include software testing to check whether a system meets its functionalities, or security testing to express what should not happen. For the latter case, fuzzing is considered as first class citizen. It consists in exercising the system with (randomly) generated and eventually modified inputs in order to test its resistance. While fuzzing is definitively the fastest and the easiest way for testing applications, it suffers from severe limitations. Indeed, the precision of the model used for input generation: a random and/or simple model cannot reach all states and significant values. Moreover, a higher model precision can result in a combinatorial explosion of test cases.

We suggest a new approach [11] whose main ingredient is to combine timing attacks with fuzzing techniques. This new approach, allows not only reducing the test space explosion, but also to simplify the fuzzing process configuration. This new testing scenario is based on observing several executions of the system and by freezing some of its parameters in order to establish a partial order on their timing evaluation. The root of our technique is to exploit timing information to classify the input data into sub-domains according to the behavior observed for specific values of the parameters. Our approach is able to discover hidden unspecified commands that may trigger computations in the tested software. Due to the specific nature of the application (the domain of the parameters is the byte) and its programming model we can also retrieve the control flow graph of

the application. The limits of the approach have been identified, and it has been tested on two applications. Validation via a coverage tool has been established.

7.1.9. System Vulnerability Analysis

Participants: Jean-Louis Lanet, Abdelhal Mesbah, Razika Lounas, Chaharezd Yayaoui.

We present in this section our effort to detect and correct some misbehaviors encountered with some firmware. We start with an attack on a secure device, such that we are able to reverse a code while the ISA is unknown and the code itself is not available. Then, we propose a formal specification of the update process of a firmware which provides the guarantee that the updated program respects the semantics of the language. In a last aspect, we try to predict the ability of a program to be attacked thanks to a Machine Learning algorithm. We demonstrated in section 7.1.8 that a state exploration is useless until the whole program is examined, we demonstrated here that approximative solutions can deal with real live programs with an affordable response time.

7.1.9.1. Reverse engineering

We believe that an adversary can gain access to different assets of the system using a black box approach.This implies of course the absence of the source code, but also sometime the absence of the binary code (romized within the soc or micro-controller, no update mechanism, no jtag, no memory extraction, no read function, and so on). In that case, the first step consists in extracting the binary code from the system. The attacker is just allowed to load data. He has then to infer enough information on the system internals and then he should be able to gain access to the native layers. In [43], we demonstrate the advantage of a graphical representation of the data in the memory can help the reverse process thanks to the abstraction provided. Our graphical tool links all the objects with a relationship based on the presence of a pointer.

In a Java based secure element, a Java application is considered as data executed by the executed program (the virtual machine) by the native processor. We introduce a first weakness in the program that allows to read an instance as an array which violate the Java type system. This weakness allows us to dump a short part of the memory which contains the meta data on a set of arrays. Thanks to this information, we generate a mimicry attack by forging pointer illegally [41]. In turns, it open the possibility to read large part of the memory as element of a forged array. Then we succeed in characterizing the memory management algorithm [12]. At the end, we transform the initial problem of finding a vulnerability in the code of a device in a black box approach to a white box problem after de-assembling the binary code.

In another work [44], we studied the byte code verification process towards an unchecked code. We found that this verification is not complete and can be bypassed. The verifier checks the semantics of the Java Card byte code. This process is split in two parts. First, the verifier loads the methods' byte code and checks the package content. For the method segment, it checks that the control flow remain inside the methods, the jump destinations are correct and so on. Secondly, for each entry point and only for these, it controls the semantics and the type correctness of the code. This step is not performed for unreachable code, while the specification states that no unreachable code should remain in the file. However, during our analysis we discovered that the verifier does some verification on the semantics of the unreachable code. Then, thanks to a fault attack (the return byte code is noped) we diverted the control flow into this unchecked area were we stored our ill-typed code leading to the execution of an aggressive shell code which in turn dumped the native layers of the card giving access to the secret key material in plain text.

7.1.9.2. Safe system update mechanism

Dynamic Software Updating (DSU) consists in updating running programs on the fly without any downtime. This feature is interesting in critical applications that must run continuously. Because updates may lead to security breaches, the question of their correctness is raised. Formal methods are a rigorous means to ensure the correctness required by applications using DSU. We propose [13] a formal verification of correctness of DSU in a Java-based embedded system. Our approach is based on three steps. First, a formal interpretation of the semantics of update operations to ensure type safety of the update. Secondly, we rely on a functional representation of byte code, the predicate transformation calculus, and a functional model of the update

mechanism to ensure the behavioral correctness of the updated programs. It is based on the use of Hoare predicate transformation to derive a specification of an updated byte code. In the last step, we use the functional representation to model the safe update point detection mechanism. This mechanism guarantees that none of the updated method active methods are active. This property is called activeness safety. We propose a functional specification that allows to derive proof obligations that guarantee the safety of the mechanism.

7.1.9.3. Prediction of system divergence

Fault attack represents one of the serious threats against embedded system security. The result of the fault injection could lead to a mutation of the code in such a way that it becomes hostile or execute a unwanted sequence of code as we demonstrated in 7.1.8 . Any successful attack may reveal a secret information stored in the card or grant an undesired authorization. We propose a methodology [5] to recognize, during the development step, the sensitive patterns to the fault attack. It is based on the concepts from text categorization and machine learning. In fact, in this method we represented the patterns using opcodes n-grams as features and we evaluated different machine learning classifiers.

In the first experiment, we evaluated all the combination of n-gram size (for n=2, n=3 and n=4), number of features using GR method to select 100, 200, ..., 500 and 1000 relevant n-grams, n-gram weighting (Term Frequency (TF), Term Frequency Inverse Document Frequency (TFIDF) and binary representations), and five classification algorithms (Naive Bayes network (NB), Decision Tree (DT), Support Vector Machine (SVM), and the boosted version of these two lasts (BDT and BSVM)) to determine the best setting. We used accuracy measure to evaluate performance of the classifiers. In addition to accuracy, we used F1, TP rate and FP rate measures to evaluate how the algorithms classified the dangerous patterns. In the first experiment, we noted that 2-gram outperformed others. Nearly 2-gram, TFIDF, 1000 features with boosted algorithm outperformed the other settings. The F1 results have shown that the classifiers are more accurate at classifying examples of the class of non dangerous pattern compared to other classes. We suggest that this might be due to the imbalance of our data set. In the second experiment, we investigated the imbalance problem. We applied SMOTE and NCR resampling techniques to overcome this class imbalance problem. We found that the outperforming setting in the resampled data set was St_{270} also with BSVM classifier. Resampled data set improves accuracy of the smallest class and keeps the accuracy of other classes.

The experimental results indicated that the resampling techniques improved the accuracy of the classifiers. In addition, our proposed method reduces the execution time of sensitive patterns classification in comparison to the mutant generator tool micro seconds instead of hours.

7.2. Results for Axis 2: Malware analysis

The detection of malicious programs is a fundamental step to be able to guarantee system security. Programs that exhibit malicious behavior, or *malware*, are commonly used in all sort of cyberattacks. They can be used to gain remote access on a system, spy on its users, exfiltrate and modify data, execute denial of services attacks, etc.

Significant efforts are being undertaken by software and data companies and researchers to protect systems, locate infections, and reverse damage inflicted by malware. Our contribution to malware analysis include the following fields:

7.2.1. Malware Detection

Participants: Axel Legay, Fabrizio Biondi, Olivier Decourbe, Mike Enescu, Thomas Given-Wilson, Annelie Heuser, Jean-Louis Lanet, Jean Quilbeuf, Alexander Zhdanov, Olivier Zendra.

Given a file or data stream, the malware detection problem consists of understanding if the file or data stream contain traces of malicious behavior. For binary executable files in particular, this requires extracting a signature of the file, so it can be compared against signatures of known clean and malicious files to determine whether the file is malicious. Binary file signatures can be divided in *syntactic* and *semantic*.

Syntactic signatures are based on properties of the file itself, like its length, hash, number and entropy of the executable and data sections, and so on. While syntactic signatures are computationally cheap to extract from binaries, it is also easy for malware creators to deploy *obfuscation* techniques that change the file's syntactic properties, hence widely mutating the signature and preventing its use for malware detection.

Semantic signatures instead are based on the binary's behavior and interactions with the system, hence are more effective at characterizing malicious files. However, they are more expensive to extract, requiring behavioral analysis and reverse-engineering of the binary. Since behavior is much harder to change than syntactic properties, against these signatures obfuscation is used to harden the file against reverse-engineering and preventing the analysis of the behavior, instead of changing it directly.

In both cases, *malware deofbuscation* is necessary to extract signatures containing actuable information that can be used to characterize the binaries as clean or malicious. Once the signatures are available, *malware classification* techniques, usually based on machine learning, are used to automatically determine whether binaries are clean or malicious starting from their signatures. Our contributions on these fields are described in the next sections.

7.2.2. Malware Deobfuscation

Participants: Axel Legay, Fabrizio Biondi, Olivier Decourbe, Mike Enescu, Thomas Given-Wilson, Annelie Heuser, Nisrine Jafri, Jean-Louis Lanet, Jean Quilbeuf.

Given a file (usually a portable executable binary or a document supporting script macros), deobfuscation refers to the preparation of the file for the purposes of further analysis. Obfuscation techniques are specifically developed by malware creators to hinder detection reverse engineering of malicious behavior. Some of these techniques include:

Packing Packing refers to the transformation of the malware code in a compressed version to be dynamically decompressed into memory and executed from there at runtime. Packing techniques are particularly effective against static analysis, since it is very difficult to determine statically the content of the unpacked memory to be executed, particularly if packing is used multiple times. The compressed code can also be encrypted, with the key being generated in a different part of the code and used by the unpacking procedure, or even transmitted remotely from a command and control (C&C) server.

Control Flow Flattening This technique aims to hinder the reconstruction of the control flow of the malware. The malware's operation are divided into basic blocks, and a dispatcher function is created that calls the blocks in the correct order to execute the malicious behavior. Each block after its execution returns control to the dispatcher, so the control flow is flattened to two levels: the dispatcher above and all the basic blocks below.

To prevent reverse engineering of the dispatcher, it is often implemented with a cryptographic hash function. A more advanced variant of this techniques embed a full virtual machine with a randomly generated instruction set, a virtual program counted, and a virtual stack in the code, and uses the machine's interpreter as the dispatcher.

Virtualization is a very effective technique to prevent reverse engineering. To contrast it, we are implementing state-of-the-art devirtualization algorithms in `angr` , allowing it to detect and ignore the virtual machine code and retrieving the obfuscated program logic. Again, we plan to contribute our improvements to the main `angr` branch, thus helping the whole security community fighting virtualized malware.

Opaque Constants and Conditionals Reversing packing and control flow flattening techniques requires understanding of the constants and conditionals in the program, hence many techniques are deployed to obfuscate them and make them unreadable by reverse engineering techniques. Such techniques are used e.g. to obfuscate the decryption keys of packed encrypted code and the conditionals in the control flow.

We have proven the efficiency of dynamic synthesis in retrieving opaque constant and conditionals, compared to the state-of-the-art approach of using SMT (Satisfiability Modulo Theories) solvers, when the input space of the opaque function is small enough. We are developing techniques based on fragmenting and analyzing by brute force the input space of opaque conditionals, and SMT constraints in general, to be integrated in SMT solvers to improve their effectiveness.

7.2.3. Malware Classification

Participants: Axel Legay, Fabrizio Biondi, Olivier Decourbe, Mike Enescu, Thomas Given-Wilson, Annelie Heuser, Nisrine Jafri, Jean-Louis Lanet, Jean Quilbeuf.

Once malicious behavior has been located, it is essential to be able to classify the malware in its specific family to know how to disinfect the system and reverse the damage inflicted on it.

While it is rare to find an actually previously unknown malware, morphic techniques are employed by malware creators to ensure that different generations of the same malware behave differently enough than it is hard to recognize them as belonging to the same family. In particular, techniques based on the syntax of the program fails against morphic malware, since syntax can be easily changed.

To this end, semantic signatures are used to classify malware in the appropriate family. Semantic signatures capture the malware's behavior, and are thus resistant to morphic and differentiation techniques that modify the malware's syntactic signatures. We are investigating semantic signatures based on the program's System Call Dependency Graph (SCDG), which have been proven to be effective and compact enough to be used in practice. SCDGs are often extracted using a technique based on pushdown automata that is ineffective against obfuscated code; instead, we are applying concolic analysis via the `angr` engine to improve speed and coverage of the extraction.

Once a semantic signature has been extracted, it has to be compared against large database of known signatures representing the various malware families to classify it. The most efficient way to obtain this is to use a supervised machine learning classifier. In this approach, the classifier is trained with a large sample of signatures malware annotated with the appropriate information about the malware families, so that it can learn to quickly and automatically classify signatures in the appropriate family. Our work on machine learning classification focuses on using SCDGs as signatures. Since SCDGs are graphs, we are investigating and adapting algorithms for the machine learning classification of graphs, usually based on measures of shared subgraphs between different graphs. One of our analysis techniques relies on common subgraph extraction, with the idea that a malicious behavior characteristic of a malware family will yield a set of common subgraphs. Another approach relies on the Weisfeiler-Lehman graph kernel which uses the presence of nodes and their neighborhoods pattern to evaluate similarity between graphs. The presence or not of a given pattern becomes a feature in a subsequent machine learning analysis through random forest or SVM.

In malware detection and classification, it is fundamental to have a false positive rate (i.e. rate of cleanware classified as malware) approaching zero, otherwise the classification system will classify hundred or thousands of cleanware files as malware, making it useless in practice. To decrease the false positive rate, the classifier is also trained with a large and representative database of cleanware, so that it can discriminate between signatures of cleanware and malware with a minimal false positive rate. We use a large database of malware and cleanware to train our classifier, thus guaranteeing a high detection rate with a small false positive rate.

We have put in place a platform for malware analysis, using dedicated hardware provided by Cisco. This platform is now fully operational and receives a daily feed of suspicious binaries for analysis. Furthermore, we developed tools for maintaining our datasets of cleanware and malware binaries, run existing syntactic analysis on them. Our toolchain is able to extract SCDGs from malwares and cleanwares and apply our classification techniques on the SCDGs.

7.2.4. Botnet Trojan Detection

Participants: Axel Legay, Fabrizio Biondi, Vesselin Bontchev, Thomas Given-Wilson, Jean Quilbeuf, Olivier Decourbe, Najah Ben Said.

Botnet trojans are a class of malware that opens a backdoor in a system and waits from further instructions from a C&C server, and possibly replicates itself somehow. A large group of systems infected by such malware is known as a botnet, and can be used by the botnet's controller to distribute spam emails (possibly carrying other malware) and perform distributed denial-of-service (DDoS) attacks. In a DDoS attack, all the systems in the botnet flood a single target with requests amounting to gigabytes or even terabytes of traffic. The target is not able to handle such traffic or to discriminate malicious request from legitimate ones, failing to provide its service.

Detecting and correctly classify botnet trojans in transit is a necessary step to be able to stop their infection. We applied our semantic classification approach on a particular family of malware, the Mirai botnet. With these experiments, we were able to confirm that the classification based on SCDG extraction and common subgraphs mining has a very low false positive rate and a high detection rate. Furthermore, our approach proved to be more accurate than detection based on syntactic signatures, without increasing the number of false positives.

7.2.5. Modular Automated Syntactic Signature Extraction (MASSE)

Participants: Axel Legay, Fabrizio Biondi, Olivier Zendra, Alexander Zhdanov, Bruno Lebon, François Déchelle.

Malware detection techniques based on syntactic signatures (or "rules") are commonly used in antivirus since their low computational cost allows them to be used on scan the files handled by the system without excessively slowing down the system. Semantic analysis techniques are relatively expensive to use, and would slow down a system significantly if used for on-access malware detection. Hence, it is common in antivirus company to use advanced semantic techniques like the SCDG-based ones we develop to detect and analyze known and unknown malware samples, and then to manually write a syntactic rule for the detection of such samples that is uploaded to the client machines.

The MASSE projects aims at providing an open-source, self-contained architecture to deploy this on a given system, company, or infrastructure, without needing to give access to the structure's data to third parties. The architecture is composed of a server executing the computationally-expensive semantic analysis, and of a number of lightweight clients performing inexpensive syntactic analysis on the client's systems. The MASSE server automatically analyzes unknown or suspicious files passing on the clients, detects the malicious ones, synthesizes syntactic signatures for them, and updates the signature databases of the clients, keeping them protected.

The MASSE server exploits modular malware analysis, supporting malware analysis modules using dynamic, static, or hybrid analysis; extracting syntactic, semantic, or hybrid signatures; using signature-based or anomaly-based detection; and any other technique the user desires, thanks to its open source malware analysis APIs. MASSE also implements pseudonymization of the signature databases, preventing an attacker to learn precisely the syntactic signatures in case some of the clients are compromised.

7.2.6. Malware IDS

Participants: Jean-Louis Lanet, Aurélien Palisse, Colas Le Guernic.

7.2.6.1. An efficient IDS for malware detection

Ransomware is a type of malware that prevents legitimate users from accessing their machine or files and demands a payment for restoring the functionalities of the infected computer. There are two classes of ransomware: the *simple lockers*, which block the usage of the computer, and *cryptors*, that encrypt files on the computer. In the case of encryption-based ransomware, the user data can only be restored with the secret key(s) used during the attack if the key is provided by the attacker.

Detecting a malware can use two options:

- The system knows the features of the malware. Features can be structural information: n-gram or graph isomorphism, or behavioral information: APIs call or system calls. Exact pattern matching algorithm or approximative algorithm (Machine learning) can be used. This approach is known as signature based and can only detect known patterns.

- The system knows its correct behavior. Any deviation of this model leads to the detection of hostile programs. This approach can detect any new attack, it does not rely on a model of the bad behavior but on the model of the correct behavior. This approach is also known as IDS (Intrusion Detection System).

In [45], [34] we apply this technique to detect malware at run time (EPS: End Point Solution). Our first solution is based on the dynamic analysis of the data transformation by the program. We propose to monitor file activity. Since it has already been proven a valid approach in terms of detection, our main goal in is to show that a good detection rate can be achieved with little to no impact on system performances. To this end, we limit our monitoring to a minimum. In order to reduce the impact on detection with a low rate of false positive, we use the chi-square goodness-of-fit test instead of Shannon entropy (*i.e.*, sensitive to compressed chunks of data). We also achieve system completeness and fine granularity by monitoring the whole file system for all userland threads. In order to evaluate our prototype implementation, Data Aware Defense (DaD), under realistic conditions, we used the bare-metal analysis platform of the LHS, Malware - O - Matic (MoM), and ran it on a large and heterogeneous (compared to the literature) live ransomware collection. We used *de facto* industry standard benchmarks to get a pertinent and reproducible assessment of the performance penalties. A second model of the correct behavior with better results has been developed (patent pending).

Our countermeasure is efficient and can be deployed on Windows 7/10 machines with a reasonable performance hit, with an average delay of 12 μs per write operation on disk, a few hundred times smaller than previous approaches. Our extensive experiments show that the more sophisticated ransomware already use mimicry attacks. However we successfully detect 99.37 % of the samples with at most 70 MB lost per sample's threads in 90% of cases and less than 7 MB in 70% of cases. Its speed and low negative rate makes it a good candidate as a first line of defense. Once a thread is deemed malicious, instead of blocking disk accesses, other more costly metrics can be used to improve the false positive rate without impacting performance, since it would not be computed for all other threads.

7.2.7. Papers

This section gathers papers that are results common to all sections above pertaining to Axis 2.

[51] (C) The largest DDoS attacks in history have been executed by devices controlled by the Mirai botnet trojan. To prevent Mirai from spreading, this paper presents and evaluates techniques to classify binary samples as Mirai based on their syntactic and semantic properties. Syntactic malware detection is shown to have a good detection rate and no false positives, but to be very easy to circumvent. Semantic malware detection is resistant to simple obfuscation and has better detection rate than syntactic detection, while keeping false positives to zero. This paper demonstrates these results, and concludes by showing how to combine syntactic and semantic analysis techniques for the detection of Mirai.

[19] (C) We present the MASSE architecture, a YARA-based open source client-server malware detection platform. MASSE includes highly effective automated syntactic malware detection rule generation for the clients based on a server-side modular malware detection system. Multiple techniques are used to make MASSE effective at detecting malware while keeping it from disrupting users and hindering reverse-engineering of its malware analysis by malware creators.

[4] (J) Control flow obfuscation techniques can be used to hinder software reverse-engineering. Symbolic analysis can counteract these techniques, but only if they can analyze obfuscated conditional statements. We evaluate the use of dynamic synthesis to complement symbolic analysis in the analysis of obfuscated conditionals. We test this approach on the taint-analysis-resistant Mixed Boolean Arithmetics (MBA) obfuscation method that is commonly used to obfuscate and randomly diversify statements. We experimentally ascertain the practical feasibility of MBA obfuscation. We study using SMT-based approaches with different state-of-the-art SMT solvers to counteract MBA obfuscation, and we show how targeted algebraic simplification can greatly reduce the analysis time. We show that synthesis-based deobfuscation is more effective than current SMT-based deobfuscation algorithms, thus proposing a synthesis-based attacker model to complement existing attacker models.

7.3. Results for Axis 3: Building a secure network stack

7.3.1. Privacy-Preserving Abuse Detection in Future Decentralised Online Social Networks

Participants: Jeffrey Burdges, Alvaro Garcia Recuero, Christian Grothoff.

Future online social networks need to not only protect sensitive data of their users, but also protect them from abusive behavior coming from malicious participants in the network. We investigated the use of supervised learning techniques to detect abusive behavior and describe privacy-preserving protocols to compute the feature set required by abuse classification algorithms in a secure and privacy-preserving way. While our method is not yet fully resilient against a strong adaptive adversary, our evaluation suggests that it will be useful to detect abusive behavior with a minimal impact on privacy.

Our results show how to combine local knowledge with private set intersection and union cardinality protocols (with masking of BLS signature to protect identity of signers/subscribers) to privately derive feature values from users in OSNs. Given an adaptive adversary that would be able to manipulate most features we propose in our supervised learning approach, it is surprising that with just three features resistant to adversarial manipulation, the algorithms still provide useful classifications.

This work was originally presented at DPM 2016 [63] and expanded upon in Álvaro García-Recuero's PhD thesis [1].

7.3.2. Fog of Trust

Participants: Jeffrey Burdges, Christian Grothoff.

The Web of Trust (WoT) used traditionally used by tools for private communication such as PGP is used to to validate individual links between participants. Using the WoT, however, leaks meta data, such that users must opt-in for it – exposing themselves to risks of privacy loss. We proposed a new method, the Fog of Trust (FoT), which uses the privacy-preserving set intersection cardinality protocol originally used in our work on abuse detection in online social networks, to support this critical step of public key verification via collaboration. In the FoT, the social relationships — which are used to verify public keys – remain hidden. This allows keys to be verified via trusted intermediaries that were established beforehand, without the need to verify each individual new contact using Trustwords. Consequently, FoT will can the same functionality as the WoT without its drawbacks to privacy.

7.3.3. Cell tower privacy

Participants: Christian Grothoff, Neal Walfield.

Context-aware applications are programs that are able to improve their performance by adapting to the current conditions, which include the user's behavior, networking conditions, and charging opportunities. In many cases, the user's location is an excellent predictor of the context. Thus, by predicting the user's future location, we can predict the future conditions. In this work, we developed techniques to identify and predict the user's location over the next 24 hours with a minimum median accuracy of 82results include our observation that cell phones sample the towers in their vicinity, which makes cell towers as-is inappropriate for use as landmarks. Motivated by this observation, we developed two techniques for processing the cell tower traces so that landmarks more closely correspond to locations, and cell tower transitions more closely correspond to user movement. We developed a prediction engine, which is based on simple sampling distributions of the form $f(t, c)$, where t is the predicted tower, and c is a set of conditions. The conditions that we considered include the time of the day, the day of the week, the current regime, and the current tower. Our family of algorithms, called TomorrowToday, achieves 89% prediction precision across all prediction trials for predictions 30 minutes in the future. This decreases slowly for predictions further in the future, and levels off for predictions approximately 4 hours in the future, at which point we achieve 82% prediction precision across all prediction trials up to 24 hours in the future. This represents a significant improvement over NextPlace, a well-cited prediction algorithm based on non-linear time series, which achieves appropriately 80% prediction precision (self reported) for predictions 30 minutes in the future, but, unlike our predictors, which try all prediction attempts, NextPlace only attempts 7% of the prediction trials on our data set [67].

7.3.4. Taler protocol improvements

Participants: Jeffrey Burdges, Florian Dold, Christian Grothoff, Marcello Stanisci.

We started modeling the Taler protocol in the framework of Provable Security, precisely defining the formal meaning of income transparency, fairness, anonymity and unforgeablity as security games. The resulting definitions and security proofs allow a more precise statement of the security of Taler in relation to the security assumptions that are being made.

The implementation of the wallet module now supports the full Taler protocol, including the refresh operation for highly efficient and privacy-preserving change.

In addition to improving the stability of the implementation of all Taler components, we added new features to the protocol that (1) allow refunds from merchants without violating privacy and (2) allow merchants to do "customer tipping", which transfers money from merchants directly to customers' wallets as a reward for doing actions on their website.

7.3.5. Mix Networking

Participants: Jeffrey Burdges, Christian Grothoff.

We have begun implementing our ratcheting scheme for providing hybrid post-quantum and forward security to the Sphinx mix network packet format. We also began collaborating with the Panoramix project and LEAP to help resolve numerous practical challenges to deploying a mix network. We shall speak about this ongoing work at the Chaos Computer Club's annual congress 34c3 in December 2017.

7.4. Other research results

7.4.1. Privacy and Security: Information-Theoretical Quantification of Security Properties

Participants: Axel Legay, Fabrizio Biondi, Olivier Zendra, Thomas Given-Wilson, Annelie Heuser, Sean Sedwards, Jean Quilbeuf, Mike Enescu.

Information theory provides a powerful quantitative approach to measuring security and privacy properties of systems. By measuring the *information leakage* of a system security properties can be quantified, validated, or falsified. When security concerns are non-binary, information theoretic measures can quantify exactly how much information is leaked. The knowledge of such information is strategic in the developments of component-based systems.

The quantitative information-theoretical approach to security models the correlation between the secret information of the system and the output that the system produces. Such output can be observed by the attacker, and the attacker tries to infer the value of the secret information by combining this information with their prior knowledge of the system.

Armed with the produced output of the system, the attacker tries to infer information about the secret information that produced the output. The quantitative analysis we consider defines and computes how much information the attacker can expect to infer (typically measured in bits). This expected leakage of bits is the information leakage of the system. This is computed by symbolically exploring the code to be analyzed, and using the symbolic constraints accumulated over the output together with a model counting algorithm to quantify the leakage.

The quantitative approach generalizes the qualitative approach and thus provides superior analysis. In particular, a system respects non-interference if and only if its leakage is equal to zero. In practice very few systems respect non-interference, and for those that don't it is imperative to be able to distinguish between the systems leaking very small amounts of secret information and systems leaking a significant amount of secret information, since only the latter are considered to pose a security vulnerability to the system.

While quantitative leakage computation is a powerful technique to detect security vulnerabilities, computing the leakage of complex programs written in low-level languages is a hard and computationally intensive task. The most common language for low-level implementation of security protocols is C, due to its efficiency, hence much of the effort in developing tools to detect vulnerabilities in source code focus on C. Recently, we have improved the state of the art in leakage quantification from C programs by proposing the usage of approximated model counting instead of precise model counting. We have shown how the approximation can improve the efficiency of leakage quantification by orders of magnitude against a logarithmic decrease in the precision of the result, often producing the same result as precise model counters much faster, and often being able to analyze cases where precise model counters would have failed. We demonstrated this technique by providing the first quantitative leakage analysis of the C code of the Heartbleed bug, showing that our technique can detect the bug in the code.

A different but equally interesting approach is followed by our new HyLEak tool. HyLeak is also able to analyze a system and compute its information leakage, i.e. the amount of information that an observer would gain by about the value of system's secret by observing its output. Contrarily to other techniques, HyLeak can analyze randomized systems, and correctly distinguish between the randomness injected in the system and the uncertainty on the secret value. This allows HyLeak to be used both on systems with explicit randomization and systems that depend on stochastic properties, like cyber-physical systems.

HyLeak uses static code analysis to divide the system to be analyzed in components. For each component, HyLeak evaluates whether it is more convenient to analyze the component using precise or statistical analysis. Each component is analyzed with the most appropriate strategy, and then the results for all components are combined together and information leakage is estimated.

The hybrid approach provides better results than both the precise and the statistical ones in terms of computation time and precision of the result. Also, it bridges the gap between cheap but imprecise statistical techniques and precise but expensive formal techniques, allowing the user to control the required precision of the result according to the computation time they have available. We evaluated HyLeak against QUAIL's precise approach and the statiatical approach implemented in the LeakWatch tool, showing that HyLeak outperforms them both. HyLeak is open source and available at https://project.inria.fr/hyleak/

Applied to shared-key cryptosystems, the information-theoretical approach allows precise reasoning about the information leakage of any secret information in the system including, the key, and the message. Recent work on max-equivocation has generalised perfect secrecy and shown the maximum achievable theoretic bounds for the security of the key and message. Achieving these theoretic maximal bounds has been proven to be achievable by Apollonian Cell Encoders (ACEs). ACEs not only allow the maximum security possible in a shared-key cryptosystem, but also allow for infinite key reuse when the key has less entropy than the message. Further, ACEs are straightforward to construct and have a compact representation making them feasible to use in practice.

Another application is to use information leakage to reason about leakage through shared resources, representing various side-channel attacks. Developmens here allow for the formalising of the leakage model through shared resources, and quantifying how significant the leakage can be. This improves on the state-of-the-art that uses only qualified leakage, and so can be precise about how much is leakage through a shared resource. Such quantification of leakage allows for scheduling of the shared resource to exploit this information to minimise leakage. Such minimisation of leakage allows for scheduling and utilisation of resources that would fail a simple quanlified test, providing solutions when prior state-of-the-art would claim impossibility. Further, a reasoned trade-off can be made between acceptable leakage and utility of the shared resource, allowing solutions that are acceptable even if not perfect.

[53] (C; submitted) Preserving privacy of private communication against an attacker is a fundamental concern of computer science security. Unconditional encryption considers the case where an attacker has unlimited computational power, hence no complexity result can be relied upon for encryption. Optimality criteria are defined for the best possible encryption over a general collection of entropy measures. This paper introduces Apollonian cell encoders, a class of shared-key cryptosystems that are proven to be universally optimal. In addition to the highest possible security for the message,

Apollonian cell encoders prove to have perfect secrecy on their key allowing unlimited key reuse. Conditions for the existence of Apollonian cell encoders are presented, as well as a constructive proof. Further, a compact representation of Apollonian cell encoders is presented, allowing for practical implementation.

[18] (C) High-security processes have to load confidential information into shared resources as part of their operation. This confidential information may be leaked (directly or indirectly) to low-security processes via the shared resource. This paper considers leakage from high-security to low-security processes from the perspective of scheduling. The workflow model is here extended to support preemption, security levels, and leakage. Formalization of leakage properties is then built upon this extended model, allowing formal reasoning about the security of schedulers. Several heuristics are presented in the form of compositional preprocessors and postprocessors as part of a more general scheduling approach. The effectiveness of such heuristics are evaluated experimentally, showing them to achieve significantly better schedulability than the state of the art. Modeling of leakage from cache attacks is presented as a case study.

[52] (C) Quantitative information flow measurement techniques have been proven to be successful in detecting leakage of confidential information from programs. Modern approaches are based on formal methods, relying on program analysis to produce a SAT formula representing the program's behavior, and model counting to measure the possible information flow. However, while program analysis scales to large codebases like the OpenSSL project, the formulas produced are too complex for analysis with precise model counting. In this paper we use the approximate model counter ApproxMC2 to quantify information flow. We show that ApproxMC2 is able to provide a large performance increase for a very small loss of precision, allowing the analysis of SAT formulas produced from complex code. We call the resulting technique ApproxFlow and test it on a large set of benchmarks against the state of the art. Finally, we show that ApproxFlow can evaluate the leakage incurred by the Heartbleed OpenSSL bug, contrarily to the state of the art.

[20] (C) We present HyLeak, a tool for reasoning about the quantity of information leakage in programs. The tool takes as input the source code of a program and analyzes it to estimate the amount of leaked information measured by mutual information. The leakage estimation is mainly based on a hybrid method that combines precise program analysis with statistical analysis using stochastic program simulation. This way, the tool combines the best of both symbolic and randomized techniques to provide more accurate estimates with cheaper analysis, in comparison with the previous tools using one of the analysis methods alone. HyLeak is publicly available and is able to evaluate the information leakage of randomized programs, even when the secret domain is large. We demonstrate with examples that HyLeaks has the best performance among the tools that are able to analyze randomized programs with similarly high precision of estimates.

[54] (J; submitted) Analysis of a probabilistic system often requires to learn the joint probability distribution of its random variables. The computation of the exact distribution is usually an exhaustive precise analysis on all executions of the system. To avoid the high computational cost of such an exhaustive search, statistical analysis has been studied to efficiently obtain approximate estimates by analyzing only a small but representative subset of the system's behavior. In this paper we propose a hybrid statistical estimation method that combines precise and statistical analyses to estimate mutual information, Shannon entropy, and conditional entropy, together with their confidence intervals. We show how to combine the analyses on different components of the system with different accuracy to obtain an estimate for the whole system. The new method performs weighted statistical analysis with different sample sizes over different components and dynamically finds their optimal sample sizes. Moreover it can reduce sample sizes by using prior knowledge about systems and a new abstraction-then-sampling technique based on qualitative analysis. To apply the method to the source code of a system, we show how to decompose the code into components and to determine the analysis method for each component by overviewing the implementation of those techniques in HyLeak tool. We demonstrate with case studies that the new method outperforms the state of the art in quantifying information leakage.

7.4.2. Security for therapeutical environments

Participants: Axel Legay, Olivier Zendra, Thomas Given-Wilson, Sean Sedwards.

This work is done in the context of the ACANTO EU project. We aim at helping develop robotic assistants to aid mobility of mobility-impaired and elderly adults. These robotic assistants provide a variety of support to their users, including: navigational assistance, social networking, social activity planning, therapeutic regime support, and diagnostic support. In Tamis, we focus on navigational assistance and social activities, as together they yield an interesting challenge in human robot interaction. The goal is to help groups of users navigate in a potentially busy dynamic environment, while also maintaining social group cohesion.

A robotic assistant has been developed before in the DALi project, acting selfishly to ensure the safe navigation of a single user. This was achieved by using the social force model and statistical model checking in a reactive planner that frequently replanned and made immediate navigational suggestions to the user. The key operational loop of this solution was to: observe the environment, model the agents in the environment in the social force model, give safety constraints for the user, and then use statistical model checking to find the optimal next move for the user.

Generalising to groups of users poses several significant difficulties. Computationally, the challenge is exponential in the number of users, considering all their possible navigational choices. Incomplete information is normal, since sensors are distributed between robotic assistants and the environment, and communication may fail, leading to different robots having different knowledge of the environment. Maintaining group cohesion is non-trivial, since group composition and position are dynamic and, unlike swarm robotics, no group member can be abandoned. Frequent replanning is necessary since there is minimal control over the users' actions, which may include ignoring the advise of the robotic assistant

The solution we designed is to abstract away from individual users in favour of groups. This refines the prior solution for a single user. Sensor information is used to obtain traces that provide behavioural information about users and pedestrians in the environment. These traces are clustered into groups that capture both location and motion behaviour. The groups are used as the social particles in the social force model, with parameters adjusted to account for group dynamics. Statistical model checking is used to find the optimal next move for the group containing the user, and the navigation for the optimal next move is displayed to the user. The effectiveness of the group abstraction mechanisms use in this refined algorithm are validated on the BIWI walking pedestrians dataset. This shows they operate correctly and effectively, even improving over human annotations, on real world data of pedestrians in a chaotic environment.

[27] (C) People with impaired physical and mental ability often find it challenging to negotiate crowded or unfamiliar environments, leading to a vicious cycle of deteriorating mobility and sociability. To address this issue the ACANTO project is developing a robotic assistant that allows its users to engage in therapeutic group social activities, building on work done in the DALi project. Key components of the ACANTO technology are social networking and group motion planning, both of which entail the sharing and broadcasting of information. Given that the system may also make use of medical records, it is clear that the issues of security, privacy, and trust are of supreme importance to ACANTO.

[58] (C; submitted) The ACANTO project is developing robotic assistants to aid the mobility and recovery of mobility-impaired and older adults. One key feature of the project's robotic assistants is aiding with navigation in chaotic environments. Prior work has solved this for a single user with a single robot, however for therapeutic outcomes ACANTO supports social groups and group activities. Thus these robotic assistants must be able to efficiently support groups of users walking together. This requires an efficient navigation solution that can handle large numbers of users, maintain (de-facto) group cohesion despite unpredictable behaviours, and operate rapidly on embedded devices. We address these challenges by: using sensor information to develop behavioural traces, clustering traces to determine groups, modeling the groups using the social force model, and finding an optimal navigation solution using statistical model checking. The new components of this solution are validated on the ETH Zürich dataset of pedestrians in an open environment.

7.4.3. Mobile air pollution sensor platform for smart-cities

Participant: Laurent Morin.

This work is organized and coordinated by the Chaire “mobilité dans une ville durable” and financed by the Foundation of Rennes 1 (https://fondation.univ-rennes1.fr/)

The purpose of this work is to design and experiment a mobile pollution sensor platform for Smart-Cities in Rennes.

The platform is integrated in the project ROAD (Rennes Open Access to Data) proposing to development of mobile systems operating the collection and the management of open data in Rennes for a future development of a smart-city. The collaboration is part of an ecosystem developed by the Chair “mobilité dans une ville durable” via the production of multiple experimentations in the city.

In the ROAD project context, the air quality in the city has been identified as one of the major challenge. Air quality improvement can only be achieved with a citizen and political full cooperation and involvement. This experimentation aims at providing an end-to-end urban platform that extends current practices in air quality measurements and allows citizens and policy makers to obtain the data and make informed decisions.

The mobile air pollution sensor platform for smart-cities proposes a innovative IoT architecture introducing the deployment of a small set of advanced and cost-effective sensors around a balanced high-performance/low-power compute unit inside a mobile agent in the city. The compute unit will have to provide the necessary computation power needed to produce advanced analysis and the security management on-site (integrity, authentication, ...).

The mobile sensor platform developments partially started in July 2017, and accelerated in October for a real deployment in buses in 2018. During this period, the core system of the platform was designed, adapted, and partially implemented to offer an operational prototype. This year lead to the design of a suitcase containing a self-sufficient measurement system: a main compute unit, its power supply and power management, and a set of satellite pollution sensors. This achievement was disseminated to the Rennes ecosystem (Rennes Atalante, Rennes Métropole, Inria) through the participation to several meetings and exhibitions.