RESEARCH CENTER

FIELD
Algorithmics, Programming, Software and Architecture

Activity Report 2018

Section New Results

Edition: 2019-03-07

Algorithmics, Computer Algebra and Cryptology

Architecture, Languages and Compilation

Embedded and Real-time Systems

Proofs and Verification

SECURITY AND CONFIDENTIALITY

ARIC Project-Team

7. New Results

7.1. Efficient approximation methods

7.1.1. A High Throughput Polynomial and Rational Function Approximations Evaluator

In [21] we present an automatic method for the evaluation of functions via polynomial or rational approximations and its hardware implementation, on FPGAs. These approximations are evaluated using Ercegovac's iterative E-method adapted for FPGA implementation. The polynomial and rational function coefficients are optimized such that they satisfy the constraints of the E-method. We present several examples of practical interest; in each case a resource-efficient approximation is proposed and comparisons are made with alternative approaches.

7.1.2. Continued fractions in power series fields

In [4], 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.

7.1.3. A Lattice Basis Reduction Approach for the Design of Finite Wordlength FIR Filters

Many applications of finite impulse response (FIR) digital filters impose strict format constraints for the filter coefficients. Such requirements increase the complexity of determining optimal designs for the problem at hand. In [6], we introduce a fast and efficient method, based on the computation of good nodes for polynomial interpolation and Euclidean lattice basis reduction. Experiments show that it returns quasi-optimal finite wordlength FIR filters; compared to previous approaches it also scales remarkably well (length 125 filters are treated in $<$ 9s). It also proves useful for accelerating the determination of optimal finite wordlength FIR filters.

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

In [7], 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 [14], 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).

7.1.6. *A Newton-like Validation Method for Chebyshev Approximate Solutions of Linear Ordinary Differential Systems*

In [22], we provide a new framework for *a posteriori* validation of vector-valued problems with componentwise tight error enclosures, and use it to design a symbolic-numeric Newton-like validation algorithm for Chebyshev approximate solutions of coupled systems of linear ordinary differential equations. More precisely, given a coupled differential system with polynomial coefficients over a compact interval (or continuous coefficients rigorously approximated by polynomials) and componentwise polynomial approximate solutions in Chebyshev basis, the algorithm outputs componentwise rigorous upper bounds for the approximation errors, with respect to the uniform norm over the interval under consideration.

A complexity analysis shows that the number of arithmetic operations needed by this algorithm (in floating-point or interval arithmetics) is proportional to the approximation degree when the differential equation is considered fixed. Finally, we illustrate the efficiency of this fully automated validation method on an example of a coupled Airy-like system.

7.1.7. *Fuel-optimal impulsive fixed-time trajectories in the linearized circular restricted 3-body-problem*

In [41], the problem of fixed-time fuel-optimal trajectories with high-thrust propulsion in the vicinity of a Lagrange point is tackled via the linear version of the primer vector theory. More precisely, the proximity to a Lagrange point i.e. any equilibrium point-stable or not-in the circular restricted three-body problem allows for a linearization of the dynamics. Furthermore, it is assumed that the spacecraft has ungimbaled thrusters, leading to a formulation of the cost function with the 1-norm for space coordinates, even though a generalization exists for steerable thrust and the 2-norm. In this context, the primer vector theory gives necessary and sufficient optimality conditions for admissible solutions to two-value boundary problems. Similarly to the case of rendezvous in the restricted two-body problem, the in-plane and out-of-plane trajectories being uncoupled, they can be treated independently. As a matter of fact, the out-of-plane dynamics is simple enough for the optimal control problem to be solved analytically via this indirect approach. As for the in-plane dynamics, the primer vector solution of the so-called primal problem is derived by solving a hierarchy of linear programs, as proposed recently for the aforementioned rendezvous. The optimal thrusting strategy is then numerically obtained from the necessary and sufficient conditions. Finally, in-plane and out-of-plane control laws are combined to form the complete 3-D fuel-optimal solution. Results are compared to the direct approach that consists in working on a discrete set of times in order to perform optimization in finite dimension. Examples are provided near various Lagrange points in the Sun-Earth and Earth-Moon systems, hinting at the extensive span of possible applications of this technique in station-keeping as well as mission analysis, for instance when connecting manifolds to achieve escape or capture.

7.2. Floating-point and Validated Numerics

7.2.1. *Optimal bounds on relative errors of floating-point operations*

Rounding error analyses of numerical algorithms are most often carried out via repeated applications of the so-called standard models of floating-point arithmetic. Given a round-to-nearest function fl and barring underflow and overflow, such models bound the relative errors $E_1(t) = |t - fl(t)|/|t|$ and $E_2(t) = |t - fl(t)|/|fl(t)|$ by the unit roundoff u. In [10] we investigate the possibility and the usefulness of refining these bounds, both in the case of an arbitrary real t and in the case where t is the exact result of an arithmetic operation on some floating-point numbers. We show that $E_1(t)$ and $E_2(t)$ are optimally bounded by $u/(1+u)$ and u, respectively, when t is real or, under mild assumptions on the base and the precision, when $t = x \pm y$ or $t = xy$ with x, y two floating-point numbers. We prove that while this remains true for division in base $\beta > 2$, smaller, attainable bounds can be derived for both division in base $\beta = 2$ and square root. This set of optimal bounds is then applied to the rounding error analysis of various numerical algorithms: in all cases, we obtain significantly shorter proofs of the best-known error bounds for such algorithms, and/or improvements on these bounds themselves.

7.2.2. On various ways to split a floating-point number

In [32] we review several ways to split a floating-point number, that is, to decompose it into the exact sum of two floating-point numbers of smaller precision. All the methods considered here involve only a few IEEE floating-point operations, with rounding to nearest and including possibly the fused multiply-add (FMA). Applications range from the implementation of integer functions such as `round` and `floor` to the computation of suitable scaling factors aimed, for example, at avoiding spurious underflows and overflows when implementing functions such as the hypotenuse.

7.2.3. Algorithms for triple-word arithmetic

Triple-word arithmetic consists in representing high-precision numbers as the unevaluated sum of three floating-point numbers. In [45], we introduce and analyze various algorithms for manipulating triple-word numbers. Our new algorithms are faster than what one would obtain by just using the usual floating-point expansion algorithms in the special case of expansions of length 3, for a comparable accuracy.

7.2.4. Error analysis of some operations involved in the Fast Fourier Transform

In [44], we are interested in obtaining error bounds for the classical FFT algorithm in floating-point arithmetic, for the 2-norm as well as for the infinity norm. For that purpose we also give some results on the relative error of the complex multiplication by a root of unity, and on the largest value that can take the real or imaginary part of one term of the FFT of a vector x, assuming that all terms of x have real and imaginary parts less than some value b.

7.3. Lattices: algorithms and cryptology

7.3.1. Reduction of orthogonal lattice bases

As a typical application, the LLL lattice basis reduction algorithm is applied to bases of the orthogonal lattice of a given integer matrix, via reducing lattice bases of a special type. With such bases in input, we have proposed in [26] a new technique for bounding from above the number of iterations required by the LLL algorithm. The main technical ingredient is a variant of the classical LLL potential, which could prove useful to understand the behavior of LLL for other families of input bases.

7.3.2. Lattice-Based Zero-Knowledge Arguments for Integer Relations

The paper [36] provides lattice-based protocols allowing to prove relations among committed integers. While the most general zero-knowledge proof techniques can handle arithmetic circuits in the lattice setting, adapting them to prove statements over the integers is non-trivial, at least if we want to handle exponentially large integers while working with a polynomial-size modulus q. For a polynomial L, the paper provides zero-knowledge arguments allowing a prover to convince a verifier that committed L-bit bitstrings x, y and z are the binary representations of integers X, Y and Z satisfying $Z = X + Y$ over $\mathbb{Z}$. The complexity of the new arguments is only linear in L. Using them, the paper constructs arguments allowing to prove inequalities $X < Z$ among committed integers, as well as arguments showing that a committed X belongs to a public interval $[\alpha, \beta]$, where α and β can be arbitrarily large. The new range arguments have logarithmic cost (i.e., linear in L) in the maximal range magnitude. Using these tools, the paper obtains zero-knowledge arguments showing that a committed element X does not belong to a public set S using $O(n \cdot \log |S|)$ bits of communication, where n is the security parameter. The paper finally gives a protocol allowing to argue that committed L-bit integers X, Y and Z satisfy multiplicative relations $Z = XY$ over the integers, with communication cost subquadratic in L. To this end, the paper uses its new protocol for integer addition to prove the correct recursive execution of Karatsuba's multiplication algorithm. The security of the new protocols relies on standard lattice assumptions with polynomial modulus and polynomial approximation factor.

7.3.3. Logarithmic-Size Ring Signatures With Tight Security from the DDH Assumption

Ring signatures make it possible for a signer to anonymously and, yet, convincingly leak a secret by signing a message while concealing his identity within a flexibly chosen ring of users. Unlike group signatures, they do not involve any setup phase or tracing authority. Despite a lot of research efforts in more than 15 years, most of their realizations require linear-size signatures in the cardinality of the ring. In the random oracle model, two recent constructions decreased the signature length to be only logarithmic in the number N of ring members. On the downside, their suffer from rather loose reductions incurred by the use of the Forking Lemma. This paper considers the problem of proving them tightly secure without affecting their space efficiency. Surprisingly, existing techniques for proving tight security in ordinary signature schemes do not trivially extend to the ring signature setting. The paper [37] overcomes these difficulties by combining the Groth-Kohlweiss Σ-protocol (Eurocrypt'15) with dual-mode encryption schemes. The main result is a fully tight construction based on the Decision Diffie-Hellman assumption in the random oracle model. By full tightness, we mean that the reduction's advantage is as large as the adversary's, up to a constant factor.

7.3.4. Adaptively Secure Distributed PRFs from LWE

In distributed pseudorandom functions (DPRFs), a PRF secret key SK is secret shared among N servers so that each server can locally compute a partial evaluation of the PRF on some input X. A combiner that collects t partial evaluations can then reconstruct the evaluation $F(SK, X)$ of the PRF under the initial secret key. So far, all non-interactive constructions in the standard model are based on lattice assumptions. One caveat is that they are only known to be secure in the static corruption setting, where the adversary chooses the servers to corrupt at the very beginning of the game, before any evaluation query. The paper [38] constructs the first fully non-interactive adaptively secure DPRF in the standard model. The construction is proved secure under the LWE assumption against adversaries that may adaptively decide which servers they want to corrupt. The new construction is also extended in order to achieve robustness against malicious adversaries.

7.3.5. Unbounded ABE via Bilinear Entropy Expansion, Revisited

This paper [24] presents simpler and improved constructions of unbounded attribute-based encryption (ABE) schemes with constant-size public parameters under static assumptions in bilinear groups. Concretely, we obtain: a simple and adaptively secure unbounded ABE scheme in composite-order groups, improving upon a previous construction of Lewko and Waters (Eurocrypt'11) which only achieves selective security; an improved adaptively secure unbounded ABE scheme based on the k-linear assumption in prime-order groups with shorter ciphertexts and secret keys than those of Okamoto and Takashima (Asiacrypt'12); the first adaptively secure unbounded ABE scheme for arithmetic branching programs under static assumptions. At the core of all of these constructions is a "bilinear entropy expansion" lemma that allows us to generate any polynomial amount of entropy starting from constant-size public parameters; the entropy can then be used to transform existing adaptively secure "bounded" ABE schemes into unbounded ones.

7.3.6. Improved Anonymous Broadcast Encryptions: Tight Security and Shorter Ciphertext

This paper [35] investigates anonymous broadcast encryptions (ANOBE) in which a ciphertext hides not only the message but also the target recipients associated with it. Following Libert et al.'s generic construction [PKC, 2012], we propose two concrete ANOBE schemes with tight reduction and better space efficiency.

- The IND-CCA security and anonymity of our two ANOBE schemes can be tightly reduced to standard k-Linear assumption (and the existence of other primitives). For a broadcast system with n users, Libert et al.'s security analysis suffers from $\mathcal{O}(n^3)$ loss while our security loss is constant.
- Our first ANOBE supports fast decryption and has a shorter ciphertext than the fast-decryption version of Libert et al.'s concrete ANOBE. Our second ANOBE is adapted from the first one.

We sacrifice the fast decryption feature and achieve shorter ciphertexts than Libert et al.'s concrete ANOBE with the help of bilinear groups. Technically, we start from an instantiation of Libert et al.'s generic ANOBE [PKC, 2012], but we work out all our proofs from scratch instead of relying on their generic security result. This intuitively allows our optimizations in the concrete setting.

7.3.7. Compact IBBE and Fuzzy IBE from Simple Assumptions

This paper [29] proposes new constructions for identity-based broadcast encryption (IBBE) and fuzzy identity-based encryption (FIBE) in composite-order groups equipped with a bilinear pairing. Our starting point is the IBBE scheme of Delerablée (Asiacrypt 2007) and the FIBE scheme of Herranz et al. (PKC 2010) proven secure under parameterized assumptions called generalized decisional bilinear Diffie-Hellman (GDDHE) and augmented multi-sequence of exponents Diffie-Hellman (aMSE-DDH) respectively. The two schemes are described in the prime-order pairing group. We transform the schemes into the setting of (symmetric) composite-order groups and prove security from two static assumptions (subgroup decision). The Déjà Q framework of Chase et al. (Asiacrypt 2016) is known to cover a large class of parameterized assumptions (dubbed "Uber assumption"), that is, these assumptions, when defined in asymmetric composite-order groups, are implied by subgroup decision assumptions in the underlying composite-order groups. We argue that the GDDHE and aMSE-DDH assumptions are not covered by the Déjà Q uber assumption framework. We therefore work out direct security reductions for the two schemes based on subgroup decision assumptions. Furthermore, our proofs involve novel extensions of Déjà Q techniques of Wee (TCC 2016-A) and Chase et al. Our constructions have constant-size ciphertexts. The IBBE has constant-size keys as well and achieves a stronger security guarantee as compared to Delerablée's IBBE, thus making it the first compact IBBE known to be selectively secure without random oracles under simple assumptions. The fuzzy IBE scheme is the first to simultaneously feature constant-size ciphertexts and security under standard assumptions.

7.3.8. Improved Inner-product Encryption with Adaptive Security and Full Attribute-hiding

This paper [25] proposes two IPE schemes achieving both adaptive security and full attribute-hiding in the prime-order bilinear group, which improve upon the unique existing result satisfying both features from Okamoto and Takashima [Eurocrypt'12] in terms of efficiency.

- Our first IPE scheme is based on the standard k-Lin assumption and has shorter master public key and shorter secret keys than Okamoto and Takashima's IPE under weaker $DLIN$=2-lin assumption.
- Our second IPE scheme is adapted from the first one; the security is based on the XDLIN assumption (as Okamoto and Takashima's IPE) but now it also enjoys shorter ciphertexts.

Technically, instead of starting from composite-order IPE and applying existing transformation, we start from an IPE scheme in a very restricted setting but already in the prime-order group, and then gradually upgrade it to our full-fledged IPE scheme. This method allows us to integrate Chen et al.'s framework [Eurocrypt'15] with recent new techniques [TCC'17, Eurocrypt'18] in an optimized way.

7.3.9. Improved Security Proofs in Lattice-Based Cryptography: Using the Rényi Divergence Rather than the Statistical Distance

The Rényi divergence is a measure of closeness of two probability distributions. In this paper [5], we show that it can often be used as an alternative to the statistical distance in security proofs for lattice-based cryptography. Using the Rényi divergence is particularly suited for security proofs of primitives in which the attacker is required to solve a search problem (e.g., forging a signature). We show that it may also be used in the case of distinguishing problems (e.g., semantic security of encryption schemes), when they enjoy a public sampleability property. The techniques lead to security proofs for schemes with smaller parameters, and sometimes to simpler security proofs than the existing ones.

7.3.10. CRYSTALS-Dilithium: A Lattice-Based Digital Signature Scheme

This paper [8] presents Dilithium, a lattice-based signature scheme that is part of the CRYSTALS (Cryptographic Suite for Algebraic Lattices) package that will be submitted to the NIST call for post-quantum standards. The scheme is designed to be simple to securely implement against side-channel attacks and to have comparable efficiency to the currently best lattice-based signature schemes. Our implementation results show that Dilithium is competitive with lattice schemes of the same security level and outperforms digital signature schemes based on other post-quantum assumptions.

7.3.11. On the asymptotic complexity of solving LWE

In this paper [9], we provide for the first time an asymptotic comparison of all known algorithms for the search version of the Learning with Errors (LWE) problem. This includes an analysis of several lattice-based approaches as well as the combinatorial BKW algorithm. Our analysis of the lattice-based approaches defines a general framework, in which the algorithms of Babai, Lindner–Peikert and several pruning strategies appear as special cases. We show that within this framework, all lattice algorithms achieve the same asymptotic complexity. For the BKW algorithm, we present a refined analysis for the case of only a polynomial number of samples via amplification, which allows for a fair comparison with lattice-based approaches. Somewhat surprisingly, such a small number of samples does not make the asymptotic complexity significantly inferior, but only affects the constant in the exponent. As the main result we obtain that both, lattice-based techniques and BKW with a polynomial number of samples, achieve running time $2^{O(n)}$ for n-dimensional LWE, where we make the constant hidden in the big-O notion explicit as a simple and easy to handle function of all LWE-parameters. In the lattice case this function also depends on the time to compute a BKZ lattice basis with block size $\Theta(n)$. Thus, from a theoretical perspective our analysis reveals how LWE 's complexity changes as a function of the LWE-parameters, and from a practical perspective our analysis is a useful tool to choose LWE-parameters resistant to all currently known attacks.

7.3.12. Measuring, Simulating and Exploiting the Head Concavity Phenomenon in BKZ

The Blockwise-Korkine-Zolotarev (BKZ) lattice reduction algorithm is central in cryptanalysis, in particular for lattice-based cryptography. A precise understanding of its practical behavior in terms of run-time and output quality is necessary for parameter selection in cryptographic design. As the provable worst-case bounds poorly reflect the practical behavior, cryptanalysts rely instead on the heuristic BKZ simulator of Chen and Nguyen (Asiacrypt'11). It fits better with practical experiments, but not entirely. In particular, it over-estimates the norm of the first few vectors in the output basis. Put differently, BKZ performs better than its Chen-Nguyen simulation.

In this article [15], we first report experiments providing more insight on this shorter-than-expected phenomenon. We then propose a refined BKZ simulator by taking the distribution of short vectors in random lattices into consideration. We report experiments suggesting that this refined simulator more accurately predicts the concrete behavior of BKZ. Furthermore, we design a new BKZ variant that exploits the shorter-than-expected phenomenon. For the same cost assigned to the underlying SVP-solver, the new BKZ variant produces bases of better quality. We further illustrate its potential impact by testing it on the SVP-120 instance of the Darmstadt lattice challenge.

7.3.13. CRYSTALS - Kyber: A CCA-Secure Module-Lattice-Based KEM

Rapid advances in quantum computing, together with the announcement by the National Institute of Standards and Technology (NIST) to define new standards for digital signature, encryption, and key-establishment protocols, have created significant interest in post-quantum cryptographic schemes. This paper [17] introduces Kyber (part of CRYSTALS - Cryptographic Suite for Algebraic Lattices - a package submitted to NIST post-quantum standardization effort in November 2017), a portfolio of post-quantum cryptographic primitives built around a key-encapsulation mechanism (KEM), based on hardness assumptions over module lattices. Our KEM is most naturally seen as a successor to the NEWHOPE KEM (Usenix 2016). In particular, the key and ciphertext sizes of our new construction are about half the size, the KEM offers CCA instead of only passive security, the security is based on a more general (and flexible) lattice problem, and our optimized implementation results in essentially the same running time as the aforementioned scheme. We first introduce a CPA-secure public-key encryption scheme, apply a variant of the Fujisaki-Okamoto transform to create a CCA-secure KEM, and eventually construct, in a black-box manner, CCA-secure encryption, key exchange, and authenticated-key-exchange schemes. The security of our primitives is based on the hardness of Module-LWE in the classical and quantum random oracle models, and our concrete parameters conservatively target more than 128 bits of postquantum security.

7.3.14. Learning with Errors and Extrapolated Dihedral Cosets

The hardness of the learning with errors (LWE) problem is one of the most fruitful resources of modern cryptography. In particular, it is one of the most prominent candidates for secure post-quantum cryptography. Understanding its quantum complexity is therefore an important goal. In this paper [20], we show that under quantum polynomial time reductions, LWE is equivalent to a relaxed version of the dihedral coset problem (DCP), which we call extrapolated DCP (eDCP). The extent of extrapolation varies with the LWE noise rate. By considering different extents of extrapolation, our result generalizes Regev's famous proof that if DCP is in BQP (quantum poly-time) then so is LWE (FOCS'02). We also discuss a connection between eDCP and Childs and Van Dam's algorithm for generalized hidden shift problems (SODA'07). Our result implies that a BQP solution for LWE might not require the full power of solving DCP, but rather only a solution for its relaxed version, eDCP, which could be easier.

7.3.15. Pairing-friendly twisted Hessian curves

This paper [27] presents efficient formulas to compute Miller doubling and Miller addition utilizing degree-3 twists on curves with j-invariant 0 written in Hessian form. We give the formulas for both odd and even embedding degrees and for pairings on both $G_1 \times G_2$ and $G_2 \times G_1$. We propose the use of embedding degrees 15 and 21 for 128-bit and 192-bit security respectively in light of the NFS attacks and their variants. We give a comprehensive comparison with other curve models; our formulas give the fastest known pairing computation for embedding degrees 15, 21, and 24.

7.3.16. On the Statistical Leak of the GGH13 Multilinear Mapand some Variants

At EUROCRYPT 2013, Garg, Gentry and Halevi proposed a candidate construction (later referred as GGH13) of cryptographic multilinear map (MMap). Despite weaknesses uncovered by Hu and Jia (EUROCRYPT 2016), this candidate is still used for designing obfuscators. The naive version of the GGH13 scheme was deemed susceptible to averaging attacks, i.e., it could suffer from a statistical leak (yet no precise attack was described). A variant was therefore devised, but it remains heuristic. Recently, to obtain MMaps with low noise and modulus, two variants of this countermeasure were developed by Döttling et al. (EPRINT:2016/599). In this work [28], we propose a systematic study of this statistical leak for all these GGH13 variants. In particular, we confirm the weakness of the naive version of GGH13. We also show that, among the two variants proposed by Döttling et al., the so-called conservative method is not so effective: it leaks the same value as the unprotected method. Luckily, the leak is more noisy than in the unprotected method, making the straightforward attack unsuccessful. Additionally, we note that all the other methods also leak values correlated with secrets. As a conclusion, we propose yet another countermeasure, for which this leak is made unrelated to all secrets. On our way, we also make explicit and tighten the hidden exponents in the size of the parameters, as an effort to assess and improve the efficiency of MMaps.

7.3.17. Higher dimensional sieving for the number field sieve algorithms

Since 2016 and the introduction of the exTNFS (extended tower number field sieve) algorithm, the security of cryptosystems based on nonprime finite fields, mainly the pairing- and torus-based ones, is being reassessed. The feasibility of the relation collection, a crucial step of the NFS variants, is especially investigated. It usually involves polynomials of degree 1, i.e., a search space of dimension 2. However, exTNFS uses bivariate polynomials of at least four coefficients. If sieving in dimension 2 is well described in the literature, sieving in higher dimensions has received significantly less attention. In this work [30], we describe and analyze three different generic algorithms to sieve in any dimension for the NFS algorithms. Our implementation shows the practicability of dimension-4 sieving, but the hardness of dimension-6 sieving.

7.3.18. Speed-Ups and Time-Memory Trade-Offs for Tuple Lattice Sieving

In this work [31], we study speed-ups and time–space trade-offs for solving the shortest vector problem (SVP) on Euclidean lattices based on tuple lattice sieving. Our results extend and improve upon previous work of Bai–Laarhoven–Stehlé [ANTS'16] and Herold–Kirshanova [PKC'17], with better complexities for arbitrary tuple sizes and offering tunable time–memory tradeoffs. The trade-offs we obtain stem from the generalization and combination of two algorithmic techniques: the configuration framework introduced by Herold–Kirshanova, and the spherical locality-sensitive filters of Becker–Ducas–Gama–Laarhoven [SODA'16]. When the available memory scales quasi-linearly with the list size, we show that with triple sieving we can solve SVP in dimension n in time $2^{0.3588n+o(n)}$ and space $2^{0.1887n+o(n)}$, improving upon the previous best triple sieve time complexity of $2^{0.3717n+o(n)}$ of Herold–Kirshanova. Using more memory we obtain better asymptotic time complexities. For instance, we obtain a triple sieve requiring only $2^{0.3300n+o(n)}$ time and $2^{0.2075n+o(n)}$ memory to solve SVP in dimension n. This improves upon the best double Gauss sieve of Becker–Ducas–Gama–Laarhoven, which runs in $2^{0.3685n+o(n)}$ time when using the same amount of space.

7.3.19. Improved Quantum Information Set Decoding

In this paper [34], we present quantum information set decoding (ISD) algorithms for binary linear codes. First, we refine the analysis of the quantum walk based algorithms proposed by Kachigar and Tillich (PQCrypto'17). This refinement allows us to improve the running time of quantum decoding in the leading order term: for an n-dimensional binary linear code the complexity of May-Meurer-Thomae ISD algorithm (Asiacrypt'11) drops down from $2^{0.05904n+o(n)}$ to $2^{0.05806n+o(n)}$. Similar improvement is achieved for our quantum version of Becker-JeuxMay-Meurer (Eurocrypt'12) decoding algorithm. Second, we translate May-Ozerov Near Neighbour technique (Eurocrypt'15) to an 'updateand-query' language more common in a similarity search literature. This re-interpretation allows us to combine Near Neighbour search with the quantum walk framework and use both techniques to improve a quantum version of Dumer's ISD algorithm: the running time goes down from $2^{0.059962n+o(n)}$ to $2^{0.059450+o(n)}$.

7.3.20. Quantum Attacks against Indistinguishablility Obfuscators Proved Secure in the Weak Multilinear Map Model

We present in [39] a quantum polynomial time attack against the GMMSSZ branching program obfuscator of Garg et al. (TCC'16), when instantiated with the GGH13 multilinear map of Garg et al. (EUROCRYPT'13). This candidate obfuscator was proved secure in the weak multilinear map model introduced by Miles et al. (CRYPTO'16). Our attack uses the short principal ideal solver of Cramer et al. (EUROCRYPT'16), to recover a secret element of the GGH13 multilinear map in quantum polynomial time. We then use this secret element to mount a (classical) polynomial time mixed-input attack against the GMMSSZ obfuscator. The main result of this article can hence be seen as a classical reduction from the security of the GMMSSZ obfuscator to the short principal ideal problem (the quantum setting is then only used to solve this problem in polynomial time). As an additional contribution, we explain how the same ideas can be adapted to mount a quantum polynomial time attack against the DGGMM obfuscator of Döttling et al. (ePrint 2016), which was also proved secure in the weak multilinear map model.

7.3.21. On the Ring-LWE and Polynomial-LWE Problems

The Ring Learning With Errors problem (RLWE) comes in various forms. Vanilla RLWE is the decision dual-RLWE variant, consisting in distinguishing from uniform a distribution depending on a secret belonging to the dual $O_K^\vee$ of the ring of integers O_K of a specified number field K. In primal-RLWE, the secret instead belongs to O_K. Both decision dual-RLWE and primal-RLWE enjoy search counterparts. Also widely used is (search/decision) Polynomial Learning With Errors (PLWE), which is not defined using a ring of integers O_K of a number field K but a polynomial ring $Z[x]/f$ for a monic irreducible $f \in Z[x]$. We show that there exist reductions between all of these six problems that incur limited parameter losses. More precisely: we prove that the (decision/search) dual to primal reduction from Lyubashevsky et al. [EUROCRYPT 2010] and Peikert [SCN 2016] can be implemented with a small error rate growth for all rings (the resulting reduction is nonuniform polynomial time); we extend it to polynomial-time reductions between (decision/search) primal

RLWE and PLWE that work for a family of polynomials f that is exponentially large as a function of deg(f) (the resulting reduction is also non-uniform polynomial time); and we exploit the recent technique from Peikert et al. [STOC 2017] to obtain a search to decision reduction for RLWE for arbitrary number fields. The reductions incur error rate increases that depend on intrinsic quantities related to K and f.

7.3.22. Non-Trivial Witness Encryption and Null-iO from Standard Assumptions

A *witness encryption (WE)* scheme can take any NP statement as a public-key and use it to encrypt a message. If the statement is true then it is possible to decrypt the message given a corresponding witness, but if the statement is false then the message is computationally hidden. Ideally, the encryption procedure should run in polynomial time, but it is also meaningful to define a weaker notion, which we call *non-trivially exponentially efficient* WE (XWE), where the encryption run-time is only required to be much smaller than the trivial 2^m bound for NP relations with witness size m. In [19], we show how to construct such XWE schemes for all of NP with encryption run-time $2^{m/2}$ under the sub-exponential learning with errors (LWE) assumption. For NP relations that can be verified in NC^1 (e.g., SAT) we can also construct such XWE schemes under the sub-exponential Decisional Bilinear Diffie-Hellman (DBDH) assumption. Although we find the result surprising, it follows via a very simple connection to *attribute-based encryption*.

We also show how to upgrade the above results to get non-trivially exponentially efficient *indistinguishability obfuscation for null circuits (niO)*, which guarantees that the obfuscations of any two circuits that always output 0 are indistinguishable. In particular, under the LWE assumptions we get a XniO scheme where the obfuscation time is $2^{n/2}$ for all circuits with input size n. It is known that in the case of indistinguishability obfuscation (iO) for all circuits, non-trivially efficient XiO schemes imply fully efficient iO schemes (Lin et al., PKC 2016) but it remains as a fascinating open problem whether any such connection exists for WE or niO.

Lastly, we explore a potential approach toward constructing fully efficient WE and niO schemes via multi-input ABE.

7.3.23. Function-Revealing Encryption

Multi-input functional encryption is a paradigm that allows an authorized user to compute a certain function—and nothing more—over multiple plaintexts given only their encryption. The particular case of two-input functional encryption has very exciting applications, including comparing the relative order of two plaintexts from their encrypted form (order-revealing encryption).

While being extensively studied, multi-input functional encryption is not ready for a practical deployment, mainly for two reasons. First, known constructions rely on heavy cryptographic tools such as multilinear maps. Second, their security is still very uncertain, as revealed by recent devastating attacks.

In [33], we investigate a simpler approach towards obtaining practical schemes for functions of particular interest. We introduce the notion of function-revealing encryption, a generalization of order-revealing encryption to any multi-input function as well as a relaxation of multi-input functional encryption. We then propose a simple construction of order-revealing encryption based on function-revealing encryption for simple functions, namely orthogonality testing and intersection cardinality. Our main result is an efficient order-revealing encryption scheme with limited leakage based on the standard DLin assumption.

7.3.24. Exploring Crypto Dark Matter: New Simple PRF Candidates and Their Applications

Pseudorandom functions (PRFs) are one of the fundamental building blocks in cryptography. We explore a new space of plausible PRF candidates that are obtained by mixing linear functions over different small moduli. Our candidates are motivated by the goals of maximizing simplicity and minimizing complexity measures that are relevant to cryptographic applications such as secure multiparty computation.

In [16], we present several concrete new PRF candidates that follow the above approach. Our main candidate is a *weak* PRF candidate (whose conjectured pseudorandomness only holds for uniformly random inputs) that first applies a secret mod-2 linear mapping to the input, and then a public mod-3 linear mapping to the result. This candidate can be implemented by depth-2 ACC^0 circuits. We also put forward a similar depth-3 *strong*

PRF candidate. Finally, we present a different weak PRF candidate that can be viewed as a deterministic variant of "Learning Parity with Noise" (LPN) where the noise is obtained via a mod-3 inner product of the input and the key.

The advantage of our approach is twofold. On the theoretical side, the simplicity of our candidates enables us to draw natural connections between their hardness and questions in complexity theory or learning theory (e.g., learnability of depth-2 ACC^0 circuits and width-3 branching programs, interpolation and property testing for sparse polynomials, and natural proof barriers for showing super-linear circuit lower bounds). On the applied side, the "piecewise-linear" structure of our candidates lends itself nicely to applications in secure multiparty computation (MPC). Using our PRF candidates, we construct protocols for distributed PRF evaluation that achieve better round complexity and/or communication complexity (often both) compared to protocols obtained by combining standard MPC protocols with PRFs like AES, LowMC, or Rasta (the latter two are specialized MPC-friendly PRFs). Our advantage over competing approaches is maximized in the setting of MPC with an honest majority, or alternatively, MPC with preprocessing.

Finally, we introduce a new primitive we call an *encoded-input PRF*, which can be viewed as an interpolation between weak PRFs and standard (strong) PRFs. As we demonstrate, an encoded-input PRF can often be used as a drop-in replacement for a strong PRF, combining the efficiency benefits of weak PRFs and the security benefits of strong PRFs. We conclude by showing that our main weak PRF candidate can plausibly be boosted to an encoded-input PRF by leveraging error-correcting codes.

7.3.25. Related-Key Security for Pseudorandom Functions Beyond the Linear Barrier

Related-key attacks (RKAs) concern the security of cryptographic primitives in the situation where the key can be manipulated by the adversary. In the RKA setting, the adversary's power is expressed through the class of related-key deriving (RKD) functions which the adversary is restricted to using when modifying keys. Bellare and Kohno (Eurocrypt 2003) first formalised RKAs and pin-pointed the foundational problem of constructing RKA-secure pseudorandom functions (RKA-PRFs). To date there are few constructions for RKA-PRFs under standard assumptions, and it is a major open problem to construct RKA-PRFs for larger classes of RKD functions. We make significant progress on this problem. In [3], we first show how to repair the Bellare-Cash framework for constructing RKA-PRFs and extend it to handle the more challenging case of classes of RKD functions that contain claws. We apply this extension to show that a variant of the NaorReingold function already considered by Bellare and Cash is an RKA-PRF for a class of affine RKD functions under the DDH assumption, albeit with an exponential-time security reduction. We then develop a second extension of the Bellare-Cash framework, and use it to show that the same Naor-Reingold variant is actually an RKA-PRF for a class of degree d polynomial RKD functions under the stronger decisional d-Diffie-Hellman inversion assumption. As a significant technical contribution, our proof of this result avoids the exponential-time security reduction that was inherent in the work of Bellare and Cash and in our first result.

7.3.26. Practical Fully Secure Unrestricted Inner Product Functional Encryption modulo p

In [23], we provide adaptively secure functional encryption schemes for the inner product functionality which are both efficient and allow for the evaluation of unbounded inner products modulo a prime p. Our constructions rely on new natural cryptographic assumptions in a cyclic group containing a subgroup where the discrete logarithm (DL) problem is easy which extend Castagnos and Laguillaumie's assumption (RSA 2015) of a DDH group with an easy DL subgroup. Instantiating our generic construction using class groups of imaginary quadratic fields gives rise to the most efficient functional encryption for inner products modulo an arbitrary large prime p. One of our schemes outperforms the DCR variant of Agrawal et al.'s protocols in terms of size of keys and ciphertexts by factors varying between 2 and 20 for a 112-bit security.

7.4. Algebraic computing and high-performance kernels

7.4.1. Generalized Hermite Reduction, Creative Telescoping and Definite Integration of D-Finite Functions

Hermite reduction is a classical algorithmic tool in symbolic integration. It is used to decompose a given rational function as a sum of a function with simple poles and the derivative of another rational function. In [18], we extend Hermite reduction to arbitrary linear differential operators instead of the pure derivative, and develop efficient algorithms for this reduction. We then apply the generalized Hermite reduction to the computation of linear operators satisfied by single definite integrals of D-finite functions of several continuous or discrete parameters. The resulting algorithm is a generalization of reduction-based methods for creative telescoping.

7.4.2. Hermite-Padé approximant bases

In [46] we design fast algorithms for the computation of approximant bases in shifted Popov normal form. For K a commutative field, let F be a matrix in $\mathsf{K}[x]^{m\times n}$ (truncated power series) and $\overrightarrow{d}$ be a degree vector, the problem is to compute a basis $P \in \mathsf{K}[x]^{m\times m}$ of the $\mathsf{K}[x]$-module of the relations $p \in \mathsf{K}[x]^{1\times m}$ such that $p(x)\cdot F(x) \equiv 0 \mod x^{\overrightarrow{d}}$. We obtain improved complexity bounds for handling arbitrary (possibly highly unbalanced) vectors $\overrightarrow{d}$. We also improve upon previously known algorithms for computing P in normalized shifted form for an arbitrary shift. Our approach combines a recent divide and conquer strategy which reduces the general case to the case where information on the output degree is available, and partial linearizations of the involved matrices.

7.4.3. Resultant of bivariate polynomials

We have proposed in [42] an algorithm for computing the resultant of two generic bivariate polynomials over a field K. For such p and q in $\mathsf{K}[x,y]$ both of degree d in x and n in y, the algorithm computes the resultant with respect to y using $(n^{2-1/\omega}d)^{1+o(1)}$ arithmetic operations, where ω is the exponent of matrix multiplication. Previous algorithms from the early 1970's required time $(n^2d)^{1+o(1)}$. We have also described some extensions of the approach to the computation of generic Gröbner bases and of characteristic polynomials of generic structured matrices and in univariate quotient algebras.

7.4.4. Recursive Combinatorial Structures: Enumeration, Probabilistic Analysis and Random Generation

The probabilistic behaviour of many data-structures, like series-parallel graphs used as a running example is this tutorial [13], can be analysed very precisely, thanks to a set of high-level tools provided by Analytic Combinatorics, as described in the book by Flajolet and Sedgewick. In this framework, recursive combinatorial definitions lead to generating function equations from which efficient algorithms can be designed for enumeration, random generation and, to some extent, asymptotic analysis. With a focus on random generation, this tutorial given at STACS first covers the basics of Analytic Combinatorics and then describes the idea of Boltzmann sampling and its realisation. The tutorial addresses a broad TCS audience and no particular pre-knowledge on analytic combinatorics is expected.

7.4.5. Linear Differential Equations as a Data-Structure

A lot of information concerning solutions of linear differential equations can be computed directly from the equation. It is therefore natural to consider these equations as a data-structure, from which mathematical properties can be computed. A variety of algorithms has thus been designed in recent years that do not aim at "solving", but at computing with this representation. Many of these results are surveyed in [11].

AROMATH Project-Team

6. New Results

6.1. Solving Polynomial Systems via Truncated Normal Forms

Participant: Bernard Mourrain.

In [12], we consider the problem of finding the isolated common roots of a set of polynomial functions defining a zero-dimensional ideal I in a ring R of polynomials over $\mathbb{C}$. We propose a general algebraic framework to find the solutions and to compute the structure of the quotient ring R/I from the null space of a Macaulay-type matrix. The affine dense, affine sparse, homogeneous, and multihomogeneous cases are treated. In the presented framework, the concept of a border basis is generalized by relaxing the conditions on the set of basis elements. This allows for algorithms to adapt the choice of basis in order to enhance the numerical stability. We present such an algorithm and show numerical results.

This is a joint work with Simon Telen and Marc Van Barel (Univ. Leuven, Belgium)

6.2. On supersolvable and nearly supersolvable line arrangements

Participant: Alexandru Dimca.

In the paper [3], we introduce a new class of line arrangements in the projective plane, called nearly supersolvable, and show that any arrangement in this class is either free or nearly free. More precisely, we show that the minimal degree of a Jacobian syzygy for the defining equation of the line arrangement, which is a subtle algebraic invariant, is determined in this case by the combinatorics. When such a line arrangement is nearly free, we discuss the splitting types and the jumping lines of the associated rank two vector bundle, as well as the corresponding jumping points, introduced recently by S. Marchesi and J. Vallès.

Joint work with Gabriel Sticlaru (Faculty of Mathematics and Informatics, Ovidius University, Romania).

6.3. Computing the monodromy and pole order filtration on Milnor fiber cohomology of plane curves

Participant: Alexandru Dimca.

In the paper [4], we describe an algorithm computing the monodromy and the pole order filtration on the Milnor fiber cohomology of any reduced projective plane curve C. The relation to the zero set of Bernstein-Sato polynomial of the defining homogeneous polynomial for C is also discussed. When C has some non weighted homogeneous singularities, then we have to assume that a conjecture holds in order to get some of our results. In all the examples computed so far this conjecture holds.

Joint work with Gabriel Sticlaru (Faculty of Mathematics and Informatics, Ovidius University, Romania).

6.4. Invariant Algebraic Sets and Symmetrization of Polynomial Systems

Participant: Evelyne Hubert.

Assuming the variety of a polynomial set is invariant under a group action, we construct, in [9], a set of invariants that cut the same variety. The construction can be seen as a generalization of the previously known construction for finite groups. The result though has to be understood outside an invariant variety which is independent of the polynomial set considered. We introduce the symmetrizations of a polynomial that are polynomials in a generating set of rational invariants. The generating set of rational invariants and the symmetrizations are constructed w.r.t. a section of the orbits of the group action.

6.5. Rational invariants of even ternary forms under the orthogonal group

Participant: Evelyne Hubert.

In [8], we determine a generating set of rational invariants of minimal cardinality for the action of the orthogonal group O_3 on the space $R[x, y, z]_{2d}$ of ternary forms of even degree $2d$. The construction relies on two key ingredients: On the one hand, the Slice Lemma allows us to reduce the problem to determining the invariants for the action on a subspace of the finite subgroup B_3 of signed permutations. On the other hand, our construction relies in a fundamental way on specific bases of harmonic polynomials. These bases provide maps with prescribed B_3-equivariance properties. Our explicit construction of these bases should be relevant well beyond the scope of this paper. The expression of the B_3-invariants can then be given in a compact form as the composition of two equivariant maps. Instead of providing (cumbersome) explicit expressions for the O_3-invariants, we provide efficient algorithms for their evaluation and rewriting. We also use the constructed B_3-invariants to determine the O_3-orbit locus and provide an algorithm for the inverse problem of finding an element in $R[x, y, z]_{2d}$ with prescribed values for its invariants. These are the computational issues relevant in brain imaging.

This is a joint work with P. Görlach (Max Planck institute, Leipzig) and T. Papadopoulo (EPI Athena, Inria SAM)

6.6. Algorithms for Computing Cubatures Based on Moment Theory

Participant: Evelyne Hubert.

Quadrature is an approximation of the definite integral of a function by a weighted sum of function values at specified points, or nodes, within the domain of integration. Gaussian quadratures are constructed to yield exact results for any polynomials of degree $2r - 1$ or less by a suitable choice of r nodes and weights. Cubature is a generalization of quadrature in higher dimension. In [2] we elaborate algorithms to compute all minimal cubatures for a given domain and a given degree. We propose first an algorithm in symbolic computation to characterize all cubatures of a given degree with a fixed number of nodes. The determination of the nodes and weights is then left to the computation of the eigenvectors of the matrix identified at the characterization stage and can be performed numerically. The characterisation of cubatures on which our algorithms are based stems from moment theory. We formulate the results there in a basis independent way : rather than considering the moment matrix, the central object in moment problems, we introduce the underlying linear map from the polynomial ring to its dual, the Hankel operator. This makes natural the use of bases of polynomials other than the monomial basis, and proves to be computationally relevant, either for numerical properties or to exploit symmetry.

Joint work with M. Collowald, (previously Université Nice Sophia Antipolis).

6.7. Products of Euclidean Metrics and Applications to Proximity Questions among Curves

Participants: Ioannis Emiris, Ioannis Psarros.

In [18], we study Approximate Nearest Neighbor (ANN) search on 1-dimensional shapes. We start with distance functions between discretized curves in Euclidean space: they appear in a wide range of applications, from road segments and molecular backbones to time-series in general dimension. For p-products of Euclidean metrics, for any positive integer p, we design simple and efficient data structures for ANN, based on randomized projections, which are of independent interest. They serve to solve proximity problems under a notion of distance between discretized curves, which generalizes both discrete Fréchet and Dynamic Time Warping distances. These are the most popular and practical approaches to comparing such curves. We offer the first data structures and query algorithms for ANN with arbitrarily good approximation factor, at the expense of increasing space usage and preprocessing time over existing methods. Query time complexity is comparable or significantly improved by our algorithms.

6.8. Efficient Random-Walk Methods for Approximating Polytope Volume

Participant: Ioannis Emiris.

In [5] we experimentally study the fundamental problem of computing the volume of a convex polytope given as an intersection of linear inequalities. We implement and evaluate practical randomized algorithms for accurately approximating the polytope's volume in high dimensions (e.g. one hundred). To carry out this efficiently we experimentally correlate the effect of parameters, such as random walk length and number of sample points, on accuracy and runtime. Moreover, we exploit the problem's geometry by implementing an iterative rounding procedure, computing partial generations of random points and designing fast polytope boundary oracles. Our publicly available code is significantly faster than exact computation and more accurate than existing approximation methods. We provide volume approximations for the Birkhoff polytopes of order 11 to 15, whereas exact methods have only computed that for order 10.

This is a joint work with Vissarion Fisikopoulos (Oracle Corp., Athens, Greece).

6.9. Randomized Embeddings with Slack and High-Dimensional Approximate Nearest Neighbor

Participants: Evangelos Anagnostopoulos, Ioannis Emiris, Ioannis Psarros.

In [1], we study the approximate nearest neighbor problem (e-ANN) in high dimensional Euclidean space with methods beyond Locality Sensitive Hashing (LSH), which has polynomial dependence in the dimension, sublinear query time, but subquadratic space requirement. In particular, we introduce a new definition of "low-quality" embeddings for metric spaces. It requires that, for some query point q, there exists an approximate nearest neighbor among the pre-images of the k approximate nearest neighbors in the target space. Focusing on Euclidean spaces, we employ random projections in order to reduce the original problem to one in a space of dimension inversely proportional to k. The k approximate nearest neighbors can be efficiently retrieved by a data structure such as BBD-trees. The same approach is applied to the problem of computing an approximate near neighbor, where we obtain a data structure requiring linear space, and query time in $O(dn^{\rho})$, for $\rho{\approx}1{-}e^2/\log{(1/e)}$. This directly implies a solution for e-ANN, while achieving a better exponent in the query time than the method based on BBD-trees. Better bounds are obtained in the case of doubling subsets of ℓ^2, by combining our method with r-nets. We implement our method in C++, and present experimental results in dimension up to 500 and 10^6 points, which show that performance is better than predicted by the analysis. In addition, we compare our ANN approach to E2LSH, which implements LSH, and we show that the theoretical advantages of each method are reflected on their actual performance.

6.10. Practical Volume Computation of Structured Convex Bodies, and an Application to Modeling Portfolio Dependencies and Financial Crises

Participants: Ioannis Emiris, Apostolos Chalkis.

In [16], we examine volume computation of general-dimensional polytopes and more general convex bodies, defined as the intersection of a simplex by a family of parallel hyperplanes, and another family of parallel hyperplanes or a family of concentric ellipsoids. Such convex bodies appear in modeling and predicting financial crises. The impact of crises on the economy (labor, income, etc.) makes its detection of prime interest for the public in general and for policy makers in particular. Certain features of dependencies in the markets clearly identify times of turmoil. We describe the relationship between asset characteristics by means of a copula; each characteristic is either a linear or quadratic form of the portfolio components, hence the copula can be constructed by computing volumes of convex bodies. We design and implement practical algorithms in the exact and approximate setting, we experimentally juxtapose them and study the tradeoff of exactness and accuracy for speed. We analyze the following methods in order of increasing generality: rejection sampling relying on uniformly sampling the simplex, which is the fastest approach, but inaccurate for small volumes; exact formulae based on the computation of integrals of probability distribution functions, which are the method of choice for intersections with a single hyperplane; an optimized Lawrence sign decomposition method, since the polytopes at hand are shown to be simple with additional structure; Markov chain Monte Carlo algorithms using random walks based on the hit-and-run paradigm generalized to nonlinear convex bodies and relying on new methods for computing a ball enclosed in the given body, such as a second-order cone program; the latter is experimentally extended to non-convex bodies with very encouraging results. Our C++ software, based on CGAL and Eigen and available on github, is shown to be very effective in up to 100 dimensions. Our results offer novel, effective means of computing portfolio dependencies and an indicator of financial crises, which is shown to correctly identify past crises. (The views expressed are those of the authors and do not necessarily reflect official positions of the European Commission.)

This is a joint work with Ludovic Calées (EU JRC, Ispra, Italy), and Vissarion Fisikopoulos (Oracle Corp., Athens, Greece).

6.11. On the maximal number of real embeddings of spatial minimally rigid graphs

Participants: Ioannis Emiris, Evangelos Bartzos.

In [15], we study the number of embeddings of minimally rigid graphs in Euclidean space R^D, which is (by definition) finite, modulo rigid transformations, for every generic choice of edge lengths. Even though various approaches have been proposed to compute it, the gap between upper and lower bounds is still enormous. Specific values and its asymptotic behavior are major and fascinating open problems in rigidity theory. Our work considers the maximal number of real embeddings of minimally rigid graphs in R^3. We modify a commonly used parametric semi-algebraic formulation that exploits the Cayley-Menger determinant to minimize the *a priori* number of complex embeddings, where the parameters correspond to edge lengths. To cope with the huge dimension of the parameter space and find specializations of the parameters that maximize the number of real embeddings, we introduce a method based on coupler curves that makes the sampling feasible for spatial minimally rigid graphs. Our methodology results in the first full classification of the number of real embeddings of graphs with 7 vertices in R^3, which was the smallest open case. Building on this and certain 8-vertex graphs, we improve the previously known general lower bound on the maximum number of real embeddings in R^3.

This is a joint work with J. Legersky (JK University, Linz, Austria) and E. Tsigaridas (PolSys, Inria).

6.12. Curved Optimal Delaunay Triangulation

Participant: Laurent Busé.

Meshes with curvilinear elements hold the appealing promise of enhanced geometric flexibility and higher-order numerical accuracy compared to their commonly-used straight-edge counterparts. However, the generation of curved meshes remains a computationally expensive endeavor with current meshing approaches: high-order parametric elements are notoriously difficult to conform to a given boundary geometry, and enforcing a smooth and non-degenerate Jacobian everywhere brings additional numerical difficulties to the meshing of complex domains. In the paper [6], we propose an extension of Optimal Delaunay Triangulations (ODT) to curved and graded isotropic meshes. By exploiting a continuum mechanics interpretation of ODT instead of the usual approximation theoretical foundations, we formulate a very robust geometry and topology optimization of Bézier meshes based on a new simple functional promoting isotropic and uniform Jacobians throughout the domain. We demonstrate that our resulting curved meshes can adapt to complex domains with high precision even for a small count of elements thanks to the added flexibility afforded by more control points and higher order basis functions.

Joint work Leman Feng (ENPC), Pierre Alliez (EPI Titane), Hervé Delingette (EPI Asclepios) and Mathieu Desbrun (CalTech, USA),

6.13. Convolution surfaces with varying radius: Formulae for skeletons made of arcs of circles and line segments

Participants: Evelyne Hubert, Alvaro Javier Fuentes Suárez.

In [19], we develop closed form formulae for the computation of the defining fields of convolutions surfaces. The formulae are obtained for power inverse kernels with skeletons made of line segments or arcs of circle. To obtain the formulae we use Creative Telescoping and describe how this technique can be used for other families of kernels and skeleton primitives. We apply the new formulae to obtain convolution surfaces around $\mathcal{G}^1$ skeletons, some of them closed curves. We showcase how the use of arcs of circles greatly improves the visualization of the surface around a general curve compared with a segment based approach.

6.14. Scaffolding a Skeleton with Quadrangular Tubes

Participant: Evelyne Hubert.

The goal of [22] is to construct a quadrilateral mesh around a one-dimensional skeleton that is as coarse as possible, the "scaffold". A skeleton allows one to quickly describe a shape, in particular a complex shape of high genus. The constructed scaffold is then a potential support for the surface representation: it provides a topology for the mesh, a domain for parametric representation (a quad mesh is ideal for tensor product splines) or, together with the skeleton, a grid support on which to project an implicit surface that is naturally defined by the skeleton through convolution. We provide a constructive algorithm to derive a quad-mesh scaffold with topologically regular cross-sections (which are also quads), and no T-junctions. We show that this construction is optimal in the sense that no coarser quad mesh with topologically regular cross-sections may be constructed. Finally, we apply an existing rotation minimization algorithm along the skeleton branches, which produces a mesh with a natural edge flow along the shape.

This is joint work with A. Panotopoulou (Dartmouth College), E. Ross (MESH consultants), K. Welker (University of Trier), G. Morin (Intitut de Recherche en Informatique de Toulouse).

6.15. Scaffolding skeletons using spherical Voronoi diagrams: feasibility, regularity and symmetry

Participants: Evelyne Hubert, Alvaro Javier Fuentes Suárez.

Given a skeleton made of line segments, in [7] we describe how to obtain a coarse quad mesh of a surface that encloses tightly the skeleton and follows its structure - the scaffold. We formalize as an Integer Linear Program the problem of constructing an optimal scaffold that minimizes the total number of quads on the mesh. We prove the feasibility of the Integer Linear Program for any skeleton. In particular we can generate these scaffolds for skeletons with cycles. We additionally show how to obtain regular scaffolds, i.e. with the same number of quad patches around each line segment, and symmetric scaffolds that respect the symmetries of the skeleton. An application to polygonization of skeleton-based implicit surfaces is also presented.

6.16. Exact conversion from Bézier tetrahedra to Bézier hexahedra

Participant: Bernard Mourrain.

Modeling and computing of trivariate parametric volumes is an important research topic in the field of three-dimensional isogeometric analysis. In [13], we propose two kinds of exact conversion approaches from Bézier tetrahedra to Bézier hexahedra with the same degree by reparametrization technique. In the first method, a Bézier tetrahedron is converted into a degenerate Bézier hexahedron, and in the second approach, a non-degenerate Bézier tetrahedron is converted into four non-degenerate Bézier hexahedra. For the proposed methods, explicit formulae are given to compute the control points of the resulting tensor-product Bézier hexahedra. Furthermore, in the second method, we prove that tetrahedral spline solids with C^k-continuity can be converted into a set of tensor-product Bézier volumes with G^k-continuity. The proposed methods can be used for the volumetric data exchange problems between different trivariate spline representations in CAD/CAE. Several experimental results are presented to show the effectiveness of the proposed methods.

This is a joint work with Gang Xu (Hanghzou, China), Yaoli Jin (Hanghzou, China), Zhoufang Xiao (Hanghzou, China), Qing Wu (Hanghzou, China), Timon Rabczuk (Weimar, Germany).

6.17. Constructing IGA-suitable planar parameterization from complex CAD boundary by domain partition and global/local optimization

Participant: Bernard Mourrain.

In the paper [14], we propose a general framework for constructing IGA-suitable planar B-spline parameterizations from given complex CAD boundaries. Instead of the computational domain bounded by four B-spline curves, planar domains with high genus and more complex boundary curves are considered. Firstly, some pre-processing operations including Bézier extraction and subdivision are performed on each boundary curve in order to generate a high-quality planar parameterization; then a robust planar domain partition framework is proposed to construct high-quality patch-meshing results with few singularities from the discrete boundary formed by connecting the end points of the resulting boundary segments. After the topology information generation of quadrilateral decomposition, the optimal placement of interior Bézier curves corresponding to the interior edges of the quadrangulation is constructed by a global optimization method to achieve a patch-partition with high quality. Finally, after the imposition of C^1/G^1-continuity constraints on the interface of neighboring Bézier patches with respect to each quad in the quadrangulation, the high-quality Bézier patch parameterization is obtained by a local optimization method to achieve uniform and orthogonal iso-parametric structures while keeping the continuity conditions between patches. The efficiency and robustness of the proposed method are demonstrated by several examples which are compared to results obtained by the skeleton-based parameterization approach.

This is a joint work with Gang Xu (Hanghzou, China), Ming Li (Zhejiang, China), Timon Rabczuk (Weimar, Germany), Jinlan Xu (Hangzhou, China), Stéphane P.A. Bordas (Luxembourg).

6.18. A Classification Approach to Efficient Global Optimization in Presence of Non-Computable Domains

Participant: Elisa Berrini.

Gaussian-Process based optimization methods have become very popular in recent years for the global optimization of complex systems with high computational costs. These methods rely on the sequential construction of a statistical surrogate model, using a training set of computed objective function values, which is refined according to a prescribed infilling strategy. However, this sequential optimization procedure can stop prematurely if the objective function cannot be computed at a proposed point. Such a situation can occur when the search space encompasses design points corresponding to an unphysical configuration, an ill-posed problem, or a non-computable problem due to the limitation of numerical solvers. To avoid such a premature stop in the optimization procedure, we propose in [11] to use a classification model to learn non-computable areas and to adapt the infilling strategy accordingly. Specifically, the proposed method splits the training set into two subsets composed of computable and non-computable points. A surrogate model for the objective function is built using the training set of computable points, only, whereas a probabilistic classification model is built using the union of the computable and non-computable training sets. The classifier is then incorporated in the surrogate-based optimization procedure to avoid proposing new points in the non-computable domain while improving the classification uncertainty if needed. The method has the advantage to automatically adapt both the surrogate of the objective function and the classifier during the iterative optimization process. Therefore, non-computable areas do not need to be a priori known. The proposed method is applied to several analytical problems presenting different types of difficulty, and to the optimization of a fully nonlinear fluid-structure interaction system. The latter problem concerns the drag minimization of a flexible hydrofoil with cavitation constraints. The efficiency of the proposed method compared favorably to a reference evolutionary algorithm, except for situations where the feasible domain is a small portion of the design space.

This is joint work with Matthieu Sacher (IRENAV), Régis Duvigneau (ACUMES), Olivier Le Maitre (LIMSI), Mathieu Durand (K-Epsilon), Frédéric Hauville (IRENAV), Jacques-André Astolfi (IRENAV).

6.19. Compressions of a polycarbonate honeycomb

Participant: André Galligo.

In [21], the in-plane compressive response of a polycarbonate honeycomb with circular close-packed cells is considered first experimentally then analytically. Under quasi-static uniaxial compression, we observed behaviors strongly depending on the orientation: for one of the two main orientations the compression is homogeneous, while for the other the deformation localizes in a very narrow band of cells. More surprisingly, for not crushing but extreme compression, when the load is released, the deformation is reversed, the localization disappears and the polycarbonate returns to its original shape. In order to explain this strange phenomena, we develop a geometric model of this honeycomb together with an expression of the bending energy. We focus on a basic mechanical element made of an elastica triangle. We also compare our description with previous experimental studies and simulations made with similar material. Finally , to illustrate mathematically this type of behavior, we present a simple model for buckling deformations with two degrees of freedom.

This is a joint work with Jean Rajchenbach (LPMC, UCA) and Bernard Rousselet (JAD, UCA).

6.20. Modeling and Computation of a liquid-vapor bubble formation

Participant: André Galligo.

The Capillary Equation correctly predicts the curvature evolution and the length of a quasi-static vapour formation. It describes a two-phase interface as a smooth curve resulting from a balance of curvatures that are influenced by surface tension and hydrostatic pressures. The present work provides insight into the application of the Capillary Equation to the prediction of single nu-cleate site phase change phenomena. In an effort to progress towards an application of the Capillary Equation to boiling events, a procedure to generating a numerical solution, in which the computational expense is reduced, is reported in [20].

This is a joint work with Frédéric Lesage (LCPI, UCA), Sebastian Minjeaud (JAD, UCA).

6.21. Axl, a geometric modeler for semi-algebraic shapes

Participants: Emmanouil Christoforou, Bernard Mourrain.

In [17], we describe the algebraic-geometric modeling platform Axl, which provides tools for the manipulation, computation and visualisation of semi-algebraic models. This includes meshes, basic geometric objects such as spheres, cylinders, cones, ellipsoids, torus, piecewise polynomial parameterisations of curves, surfaces or volumes such as B-spline parameterisations, as well as algebraic curves and surfaces defined by polynomial equations. Moreover, Axl provides algorithms for processing these geometric representations, such as computing intersection loci (points, curves) of parametric models, singularities of algebraic curves or surfaces, certified topology of curves and surfaces, etc. We present its main features and describe its generic extension mechanism, which allows one to define new data types and new processes on the data, which benefit from automatic visualisation and interaction facilities. The application capacities of the software are illustrated by short descriptions of plugins on algebraic curves and surfaces and on splines for Isogeometric Analysis.

This is a joint work with Angelos Mantzaflaris (JKU, Austria), Julien Wintz (SED, Inria).

CARAMBA Project-Team

7. New Results

7.1. A new family of pairing-friendly elliptic curves

Participant: Aurore Guillevic.

In [11], together with M. Scott from Miracl, we presented an algorithm to generate new families of pairing-friendly curves. It generalizes the very popular Barreto-Naehrig curves. This paper jointly received the best paper award of the conference.

7.2. Faster individual discrete logarithms in finite fields of composite extension degree

Participant: Aurore Guillevic.

We improved in [7] the previous work [25] on speeding-up the first phase of the individual discrete logarithm computation, the initial splitting, a.k.a. the smoothing phase. We extended the algorithm to any non-prime finite field $\mathbb{F}_{p^n}$ where n is composite. We also applied it to the new variant Tower-NFS. The paper is now published.

7.3. Polynomial Time Bounded Distance Decoding near Minkowski's Bound in Discrete Logarithm Lattices

Participant: Cécile Pierrot [contact].

In [6], together with Léo Ducas, we proposed a concrete family of dense lattices of arbitrary dimension n in which the lattice Bounded Distance Decoding (BDD) problem can be solved in deterministic polynomial time. The lattice construction needs discrete logarithm computations that can be made in deterministic polynomial time for well-chosen parameters. Each lattice comes with a deterministic polynomial time decoding algorithm able to decode up to a large radius. Namely, we reached decoding radius within $O(\log n)$ Minkowski's bound, for both ℓ_1 and ℓ_2-norms.

7.4. Improved complexity bounds for counting points on hyperelliptic curves

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

In [3], we presented 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.

7.5. Counting points on genus-3 hyperelliptic curves with explicit real multiplication

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

In [9], we proposed a Las Vegas probabilistic algorithm to compute the zeta function of a genus-3 hyperelliptic curve defined over a finite field $\mathbb{F}_q$, with explicit real multiplication by an order $\mathbb{Z}[\eta]$ in a totally real cubic field. Our main result states that this algorithm requires an expected number of $O((\log q)^6)$ bit-operations, where the constant in the $O()$ depends on the ring $\mathbb{Z}[\eta]$ and on the degrees of polynomials representing the endomorphism η. As a proof-of-concept, we computed the zeta function of a curve defined over a 64-bit prime field, with explicit real multiplication by $\mathbb{Z}[2\cos(2\pi/7)]$.

7.6. Counting points on hyperelliptic curves with explicit real multiplication in arbitrary genus

Participant: Simon Abelard.

In [14], we presented a probabilistic Las Vegas algorithm for computing the local zeta function of a genus-g hyperelliptic curve defined over $\mathbb{F}_q$ with explicit real multiplication (RM) by an order $\mathbb{Z}[\eta]$ in a degree-g totally real number field. It is based on the approaches by Schoof and Pila in a more favorable case where we can split the ℓ-torsion into g kernels of endomorphisms, as introduced by Gaudry, Kohel, and Smith in genus 2. To deal with these kernels in any genus, we adapted a technique that Abelard, Gaudry, and Spaenlehauer introduced to model the ℓ-torsion by structured polynomial systems. Applying this technique to the kernels, the systems we obtained are much smaller and so is the complexity of solving them. Our main result is that there exists a constant $c > 0$ such that, for any fixed g, this algorithm has expected time and space complexity $O((\log q)^c)$ as q grows and the characteristic is large enough. We proved that $c \leq 8$ and we also conjecture that the result still holds for $c = 6$.

7.7. A fast randomized geometric algorithm for computing Riemann-Roch spaces

Participants: Aude Le Gluher, Pierre-Jean Spaenlehauer [contact].

In [16], we proposed a probabilistic Las Vegas variant of Brill-Noether's algorithm for computing a basis of the Riemann-Roch space $L(D)$ associated to a divisor D on a projective plane curve $\mathcal{C}$ over a sufficiently large perfect field k. Our main result shows that this algorithm requires at most $O(\max(\deg(\mathcal{C})^{2\omega}, \deg(D_+)^{\omega}))$ arithmetic operations in k, where ω is a feasible exponent for matrix multiplication and D_+ is the smallest effective divisor such that $D_+ \geq D$. This improves the best known upper bounds on the complexity of computing Riemann-Roch spaces. Our algorithm may fail, but we showed that provided that a few mild assumptions are satisfied, the failure probability is bounded by $O(\max(\deg(\mathcal{C})^4, \deg(D_+)^2)/|E|)$, where E is a finite subset of k in which we pick elements uniformly at random. We provide a freely available C++/NTL implementation of the proposed algorithm, and experimental data. In particular, our implementation enjoys a speed-up larger than 9 on several examples compared to the reference implementation in the Magma computer algebra system. As a by-product, our algorithm also yields a method for computing the group law on the Jacobian of a smooth plane curve of genus g within $O(g^{\omega})$ operations in k, which slightly improves in this context the best known complexity $O(g^{\omega+\varepsilon})$ of Khuri-Makdisi's algorithm.

7.8. Formal proof of mpfr_add

Participants: Jianyang Pan, Paul Zimmermann [contact].

With the help of Karthik Bhargavan (Prosecco project-team), we proved formally the correctness of the `mpfr_add` code in case where all inputs and the output have the same precision, and this precision is less than one limb (i.e., less than 64 bits on modern computers). The algorithm was proven formally correct using the F^* language, and the extracted code, which was shown to be as efficient as the original MPFR code, is now available in MPFR. A similar work was done for the multiplication `mpfr_mul`, but the proof of correctness was only partly completed.

7.9. Various ways to split a floating-point number

Participant: Paul Zimmermann.

Together with Claude-Pierre Jeannerod and Jean-Michel Muller (AriC project-team), we revisited in an unified way the classical algorithms to split a floating-point number in two parts, and some applications of these algorithms. Some new algorithms were also designed. This work was presented at the 25th IEEE Symposium on Computer Arithmetic [10].

7.10. A polyhedral method for sparse systems with many positive solutions

Participant: Pierre-Jean Spaenlehauer.

Together with Frédéric Bihan (Université Savoie Mont Blanc) and Francisco Santos (Universidad de Cantabria), we investigated in [4] a version of Viro's method for constructing polynomial systems with many positive solutions, based on regular triangulations of the Newton polytope of the system. The number of positive solutions obtained with our method is governed by the size of the largest positively decorable subcomplex of the triangulation. Here, positive decorability is a property that we introduced and which is dual to being a subcomplex of some regular triangulation. Using this duality, we produced large positively decorable subcomplexes of the boundary complexes of cyclic polytopes. As a byproduct we obtained new lower bounds, some of them being the best currently known, for the maximal number of positive solutions of polynomial systems with prescribed numbers of monomials and variables. We also studied the asymptotics of these numbers and observed a log-concavity property.

7.11. Fast Integer Multiplication Using Generalized Fermat Primes

Participants: Svyatoslav Covanov, Emmanuel Thomé [contact].

In [5] we described an algorithm for the multiplication of two n-bit integers. It achieves the best asymptotic complexity bound $O(n \log n \cdot 4^{\log^* n})$ under a hypothesis on the distribution of generalized Fermat primes of the form $r^{2^\lambda} + 1$. This hypothesis states that there always exists a sufficiently small interval in which we can find such a prime. Experimental results support this assumption. This article was submitted to Mathematics of Computation and was completely rewritten in late 2017-early 2018. It is now accepted for final publication.

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

Participant: Svyatoslav Covanov.

In [15], we described a method improving on the exhaustive search algorithm originally developed in [19]. 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 proved 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. This work has been submitted to *Theoretical Computer Science* and is tentatively accepted, pending minor revisions.

7.13. Using Constraint Programming to Solve a Cryptanalytic Problem

Participant: Marine Minier.

In [8], we described Constraint Programming (CP) models to solve a cryptanalytic problem: the related key differential attacks against the standard block cipher AES. We improved our models for those attacks and the time required to solve the related key differential attacks for all instances of this particular problem. In particular, we were able to find the best related key differential trails for all the instances of AES-128, AES-192 and AES-256 in less than 5 core-hours except for one instance (AES-128 with 5 rounds) that took 15 core-hours.

7.14. Preparation of a submission for the NIST call dedicated to standardization of lightweight cryptography

Participants: Marine Minier [contact], Paul Huynh, Virginie Lallemand.

During these last six months, we prepared a submission to the NIST call dedicated on lightweight cryptography. The criteria required by this call are various and concern both small embedded micro-controllers and efficient hardware implementation with side channel and fault attack resistance. The proposal will be submitted by the call deadline, at the latest on Feb 25th, 2019.

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):

- Advanced primitives for privacy in the cloud
- Efficient functional encryption
- Several predicate-encryption schemes
- New primitives for efficient anonymous authentication
- Analyses of currently deployed zero-knowledge SNARKs

DATASHAPE Project-Team

7. New Results

7.1. Algorithmic aspects of topological and geometric data analysis

7.1.1. DTM-based filtrations

Participants: Frédéric Chazal, Marc Glisse, Raphaël Tinarrage.

In collaboration with H. Anai, Y. Ike, H. Inakoshi and Y. Umeda of Fujitsu.

Despite strong stability properties, the persistent homology of filtrations classically used in Topological Data Analysis, such as, e.g. the Čech or Vietoris-Rips filtrations, are very sensitive to the presence of outliers in the data from which they are computed. In this paper [33], we introduce and study a new family of filtrations, the DTM-filtrations, built on top of point clouds in the Euclidean space which are more robust to noise and outliers. The approach adopted in this work relies on the notion of distance-to-measure functions, and extends some previous work on the approximation of such functions.

7.1.2. Persistent Homology with Dimensionality Reduction: k-Distance vs Gaussian Kernels

Participants: Shreya Arya, Jean-Daniel Boissonnat, Kunal Dutta.

We investigate the effectiveness of dimensionality reduction for computing the persistent homology for both k-distance and kernel distance [34]. For k-distance, we show that the standard Johnson-Lindenstrauss reduction preserves the k-distance, which preserves the persistent homology upto a $(1-\varepsilon)^{-1}$ factor with target dimension $O(k\log n/\varepsilon^2)$. We also prove a concentration inequality for sums of dependent chi-squared random variables, which, under some conditions, allows the persistent homology to be preserved in $O(\log n/\varepsilon^2)$ dimensions. This answers an open question of Sheehy. For Gaussian kernels, we show that the standard Johnson-Lindenstrauss reduction preserves the persistent homology up to an $4(1-\epsilon)^{-1}$ factor.

7.1.3. Computing Persistent Homology of Flag Complexes via Strong Collapses

Participants: Jean-Daniel Boissonnat, Siddharth Pritam.

In collaboration with Divyansh Pareek (Indian Institute of Technology Bombay, India)

We introduce a fast and memory efficient approach to compute the persistent homology (PH) of a sequence of simplicial complexes. The basic idea is to simplify the complexes of the input sequence by using strong collapses, as introduced by J. Barmak and E. Miniam [DCG (2012)], and to compute the PH of an induced sequence of reduced simplicial complexes that has the same PH as the initial one. Our approach has several salient features that distinguishes it from previous work. It is not limited to filtrations (i.e. sequences of nested simplicial subcomplexes) but works for other types of sequences like towers and zigzags. To strong collapse a simplicial complex, we only need to store the maximal simplices of the complex, not the full set of all its simplices, which saves a lot of space and time. Moreover, the complexes in the sequence can be strong collapsed independently and in parallel. Finally, we can compromize between precision and time by choosing the number of simplicial complexes of the sequence we strong collapse. As a result and as demonstrated by numerous experiments on publicly available data sets, our approach is extremely fast and memory efficient in practice [27].

7.1.4. Strong Collapse for Persistence

Participants: Jean-Daniel Boissonnat, Siddharth Pritam.

In this paper, we build on the initial success of and show that further decisive progress can be obtained if one restricts the family of simplicial complexes to flag complexes. Flag complexes are fully characterized by their graph (or 1-skeleton), the other faces being obtained by computing the cliques of the graph. Hence, a flag complex can be represented by its graph, which is a very compact representation. Flag complexes are very popular and, in particular, Vietoris-Rips complexes are by far the most widely simplicial complexes used in Topological Data Analysis. It has been shown in that the persistent homology of Vietoris-Rips filtrations can be computed very efficiently using strong collapses. However, most of the time was devoted to computing the maximal cliques of the complex prior to their strong collapse. In this paper [37], we observe that the reduced complex obtained by strong collapsing a flag complex is itself a flag complex. Moreover, this reduced complex can be computed using only the 1-skeleton (or graph) of the complex, not the set of its maximal cliques. Finally, we show how to compute the equivalent filtration of the sequence of reduced flag simplicial complexes using again only 1-skeletons. x On the theory side, we show that strong collapses of flag complexes can be computed in time $O(v^2k^2)$ where v is the number of vertices of the complex and k the maximal degree of its graph. The algorithm described in this paper has been implemented and the code will be soon released in the Gudhi library. Numerous experiments show that our method outperforms previous methods, e.g. Ripser.

7.1.5. Triangulating submanifolds: An elementary and quantified version of Whitney's method

Participants: Jean-Daniel Boissonnat, Siargey Kachanovich, Mathijs Wintraecken.

We quantize Whitney's construction to prove the existence of a triangulation for any C^2 manifold, so that we get an algorithm with explicit bounds. We also give a new elementary proof, which is completely geometric [36].

7.1.6. Randomized incremental construction of Delaunay triangulations of nice point sets

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

In collaboration with Olivier Devillers (Inria, CNRS, Loria, Université de Lorraine).

Randomized incremental construction (RIC) is one of the most important paradigms for building geometric data structures. Clarkson and Shor developed a general theory that led to numerous algorithms that are both simple and efficient in theory and in practice.

Randomized incremental constructions are most of the time space and time optimal in the worst-case, as exemplified by the construction of convex hulls, Delaunay triangulations and arrangements of line segments.

However, the worst-case scenario occurs rarely in practice and we would like to understand how RIC behaves when the input is nice in the sense that the associated output is significantly smaller than in the worst-case. For example, it is known that the Delaunay triangulations of nicely distributed points in $\mathbb{R}^d$ or on polyhedral surfaces in $\mathbb{R}^3$ has linear complexity, as opposed to a worst-case complexity of $\Theta(n^{\lfloor d/2 \rfloor})$ in the first case and quadratic in the second. The standard analysis does not provide accurate bounds on the complexity of such cases and we aim at establishing such bounds in this paper [35]. More precisely, we will show that, in the two cases above and variants of them, the complexity of the usual RIC is $O(n \log n)$, which is optimal. In other words, without any modification, RIC nicely adapts to good cases of practical value.

Along the way, we prove a probabilistic lemma for sampling without replacement, which may be of independent interest.

7.1.7. Approximate Polytope Membership Queries

Participant: Guilherme Da Fonseca.

In collaboration with Sunil Arya (Hong Kong University of Science and Technology) and David Mount (University of Maryland).

In the polytope membership problem, a convex polytope K in $\mathbb{R}^d$ is given, and the objective is to preprocess K into a data structure so that, given any query point $q \in \mathbb{R}^d$, it is possible to determine efficiently whether $q \in K$. We consider this problem in an approximate setting. Given an approximation parameter ϵ, the query can be answered either way if the distance from q to K's boundary is at most ϵ times K's diameter. We assume that the dimension d is fixed, and K is presented as the intersection of n halfspaces. Previous solutions to approximate polytope membership were based on straightforward applications of classic polytope approximation techniques by Dudley (1974) and Bentley et al. (1982). The former is optimal in the worst-case with respect to space, and the latter is optimal with respect to query time. We present four main results. First, we show how to combine the two above techniques to obtain a simple space-time trade-off. Second, we present an algorithm that dramatically improves this trade-off. In particular, for any constant $\alpha \geq 4$, this data structure achieves query time roughly $O(1/\epsilon^{(d-1)/\alpha})$ and space roughly $O(1/\epsilon^{(d-1)(1-\Omega(\log \alpha))/\alpha})$. We do not know whether this space bound is tight, but our third result shows that there is a convex body such that our algorithm achieves a space of at least $\Omega(1/\epsilon^{(d-1)(1-O(\sqrt{\alpha}))/\alpha})$. Our fourth result shows that it is possible to reduce approximate Euclidean nearest neighbor searching to approximate polytope membership queries. Combined with the above results, this provides significant improvements to the best known space-time trade-offs for approximate nearest neighbor searching in $\mathbb{R}^d$. For example, we show that it is possible to achieve a query time of roughly $O(\log n + 1/\epsilon^{d/4})$ with space roughly $O(n/\epsilon^{d/4})$, thus reducing by half the exponent in the space bound [11].

7.1.8. Approximate Convex Intersection Detection with Applications to Width and Minkowski Sums

Participant: Guilherme Da Fonseca.

In collaboration with Sunil Arya (Hong Kong University of Science and Technology) and David Mount (University of Maryland).

Approximation problems involving a single convex body in d-dimensional space have received a great deal of attention in the computational geometry community. In contrast, works involving multiple convex bodies are generally limited to dimensions $d \leq 3$ and/or do not consider approximation. In this paper, we consider approximations to two natural problems involving multiple convex bodies: detecting whether two polytopes intersect and computing their Minkowski sum. Given an approximation parameter $\epsilon > 0$, we show how to independently preprocess two polytopes A, B into data structures of size $O(1/\epsilon^{(d-1)/2})$ such that we can answer in polylogarithmic time whether A and B intersect approximately. More generally, we can answer this for the images of A and B under affine transformations. Next, we show how to ϵ-approximate the Minkowski sum of two given polytopes defined as the intersection of n halfspaces in $O(n \log(1/\epsilon) + 1/\epsilon^{(d-1)/2+\alpha})$ time, for any constant $\alpha > 0$. Finally, we present a surprising impact of these results to a well studied problem that considers a single convex body. We show how to ϵ-approximate the width of a set of n points in $O(n \log(1/\epsilon) + 1/\epsilon^{(d-1)/2+\alpha})$ time, for any constant $\alpha > 0$, a major improvement over the previous bound of roughly $O(n + 1/\epsilon^{d-1})$ time [22].

7.1.9. Approximating the Spectrum of a Graph

Participant: David Cohen-Steiner.

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

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, $\lambda = (\lambda_1, \cdots, \lambda_{|V|})$, 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 $\widetilde{\lambda} = (\widetilde{\lambda}_1, \cdots, \widetilde{\lambda}_{|V|})$, such that $\|\widetilde{\lambda} - \lambda\|_1 \leq \varepsilon|V|$. Our algorithm has query complexity and running time $\exp(O(1/\varepsilon))$, which is independent of the size of the graph, $|V|$. We demonstrate the practical viability of our algorithm on synthetically generated graphs, and 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. The spectra of these real-world networks reveal insights into the structural similarities and differences between them, illustrating the potential value of our algorithm for efficiently approximating the spectrum of large networks [29].

7.1.10. Spectral Properties of Radial Kernels and Clustering in High Dimensions

Participants: David Cohen-Steiner, Alba Chiara de Vitis.

In this paper [40], we study the spectrum and the eigenvectors of radial kernels for mixtures of distributions in $\mathbb{R}^n$. Our approach focuses on high dimensions and relies solely on the concentration properties of the components in the mixture. We give several results describing of the structure of kernel matrices for a sample drawn from such a mixture. Based on these results, we analyze the ability of kernel PCA to cluster high dimensional mixtures. In particular, we exhibit a specific kernel leading to a simple spectral algorithm for clustering mixtures with possibly common means but different covariance matrices. This algorithm will succeed if the angle between any two covariance matrices in the mixture (seen as vectors in $\mathbb{R}^{n^2}$) is larger than $\Omega(n^{-1/6}\log^{5/3} n)$. In particular, the required angular separation tends to 0 as the dimension tends to infinity. To the best of our knowledge, this is the first polynomial time algorithm for clustering such mixtures beyond the Gaussian case.

7.1.11. Exact computation of the matching distance on 2-parameter persistence modules

Participant: Steve Oudot.

In collaboration with Michael Kerber (T.U. Graz) and Michael Lesnick (SUNY).

The matching distance is a pseudometric on multi-parameter persistence modules, defined in terms of the weighted bottleneck distance on the restriction of the modules to affine lines. It is known that this distance is stable in a reasonable sense, and can be efficiently approximated, which makes it a promising tool for practical applications. In [44] we show that in the 2-parameter setting, the matching distance can be computed exactly in polynomial time. Our approach subdivides the space of affine lines into regions, via a line arrangement. In each region, the matching distance restricts to a simple analytic function, whose maximum is easily computed. As a byproduct, our analysis establishes that the matching distance is a rational number, if the bigrades of the input modules are rational.

7.1.12. A Comparison Framework for Interleaved Persistence Modules

Participant: Miroslav Kramár.

In collaboration with Rachel Levanger (UPenn), Shaun Harker and Konstantin Mischaikow (Rutgers).

In [43], we present a generalization of the induced matching theorem of [1] and use it to prove a generalization of the algebraic stability theorem for R-indexed pointwise finite-dimensional persistence modules. Via numerous examples, we show how the generalized algebraic stability theorem enables the computation of rigorous error bounds in the space of persistence diagrams that go beyond the typical formulation in terms of bottleneck (or log bottleneck) distance.

7.1.13. Discrete Morse Theory for Computing Zigzag Persistence

Participant: Clément Maria.

In collaboration with Hannah Schreiber (Graz University of Technology, Austria)

We introduce a framework to simplify zigzag filtrations of general complexes using discrete Morse theory, in order to accelerate the computation of zigzag persistence. Zigzag persistence is a powerful algebraic generalization of persistent homology. However, its computation is much slower in practice, and the usual optimization techniques cannot be used to compute it. Our approach is different in that it preprocesses the filtration before computation. Using discrete Morse theory, we get a much smaller zigzag filtration with same persistence. The new filtration contains general complexes. We introduce new update procedures to modify on the fly the algebraic data (the zigzag persistence matrix) under the new combinatorial changes induced by the Morse reduction. Our approach is significantly faster in practice [45].

7.2. Statistical aspects of topological and geometric data analysis

7.2.1. Robust Bregman Clustering

Participants: Claire Brécheteau, Clément Levrard.

In collaboration with Aurélie Fischer (Université Paris-Diderot).

Using a trimming approach, in [38], we investigate a k-means type method based on Bregman divergences for clustering data possibly corrupted with clutter noise. The main interest of Bregman divergences is that the standard Lloyd algorithm adapts to these distortion measures, and they are well-suited for clustering data sampled according to mixture models from exponential families. We prove that there exists an optimal codebook, and that an empirically optimal codebook converges a.s. to an optimal codebook in the distortion sense. Moreover, we obtain the sub-Gaussian rate of convergence for k-means 1 √ n under mild tail assumptions. Also, we derive a Lloyd-type algorithm with a trimming parameter that can be selected from data according to some heuristic, and present some experimental results.

7.2.2. Statistical analysis and parameter selection for Mapper

Participants: Mathieu Carrière, Bertrand Michel, Steve Oudot.

In [15] 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.

7.2.3. A Fuzzy Clustering Algorithm for the Mode-Seeking Framework

Participants: Thomas Bonis, Steve Oudot.

In [13] we propose a new soft clustering algorithm based on the mode-seeking framework. Given a point cloud in $\mathbb{R}^d$, we define regions of high density that we call cluster cores, then we implement a random walk on a neighborhood graph built on top of the data points. This random walk is designed in such a way that it is attracted by high-density regions, the intensity of the attraction being controlled by a temperature parameter $\beta > 0$. The membership of a point to a given cluster is then the probability for the random walk starting at this point to hit the corresponding cluster core before any other. While many properties of random walks (such as hitting times, commute distances, etc) are known to eventually encode purely local information when the number of data points grows to infinity, the regularization introduced by the use of cluster cores allows the output of our algorithm to converge to quantities involving the global structure of the underlying density function. Empirically, we show how the choice of β influences the behavior of our algorithm: for small values of β the result is really close to hard mode-seeking, while for values of β close to 1 the result is similar to the output of the (soft) spectral clustering. We also demonstrate the scalability of our approach experimentally.

7.2.4. Large Scale computation of Means and Clusters for Persistence Diagrams using Optimal Transport

Participants: Théo Lacombe, Steve Oudot.

In collaboration with Marco Cuturi (ENSAE).

Persistence diagrams (PDs) are at the core of topological data analysis. They provide succinct descriptors encoding the underlying topology of sophisticated data. PDs are backed-up by strong theoretical results regarding their stability and have been used in various learning contexts. However, they do not live in a space naturally endowed with a Hilbert structure where natural metrics are not even differentiable, thus not suited to optimization process. Therefore, basic statistical notions such as the barycenter of a finite sample of PDs are not properly defined. In [30] we provide a theoretically good and computationally tractable framework to estimate the barycenter of a set of persistence diagrams. This construction is based on the theory of Optimal Transport (OT) and endows the space of PDs with a metric inspired from regularized Wasserstein distances.

7.2.5. The k-PDTM : a coreset for robust geometric inference

Participants: Claire Brécheteau, Clément Levrard.

Analyzing the sub-level sets of the distance to a compact sub-manifold of $\mathbb{R}^d$ is a common method in TDA to understand its topology. The distance to measure (DTM) was introduced by Chazal, Cohen-Steiner and Mérigot to face the non-robustness of the distance to a compact set to noise and outliers. This function makes possible the inference of the topology of a compact subset of $\mathbb{R}^d$ from a noisy cloud of n points lying nearby in the Wasserstein sense. In practice, these sub-level sets may be computed using approximations of the DTM such as the q-witnessed distance or other power distance. These approaches lead eventually to compute the homology of unions of n growing balls, that might become intractable whenever n is large. To simultaneously face the two problems of large number of points and noise, we introduce in [39] the k-power distance to measure (k-PDTM). This new approximation of the distance to measure may be thought of as a k-coreset based approximation of the DTM. Its sublevel sets consist in union of k-balls, $k << n$, and this distance is also proved robust to noise. We assess the quality of this approximation for k possibly dramatically smaller than n, for instance $k = n13$ is proved to be optimal for 2-dimensional shapes. We also provide an algorithm to compute this k-PDTM.

7.2.6. The density of expected persistence diagrams and its kernel based estimation

Participants: Frédéric Chazal, Vincent Divol.

Persistence diagrams play a fundamental role in Topological Data Analysis where they are used as topological descriptors of filtrations built on top of data. They consist in discrete multisets of points in the plane $\mathbb{R}^2$ that can equivalently be seen as discrete measures in $\mathbb{R}^2$. When the data come as a random point cloud, these discrete measures become random measures whose expectation is studied in this paper. In [28] we first show that for a wide class of filtrations, including the Čech and Rips-Vietoris filtrations, the expected persistence diagram, that is a deterministic measure on $\mathbb{R}^2$, has a density with respect to the Lebesgue measure. Second, building on the previous result we show that the persistence surface recently introduced by Adams et al can be seen as a kernel estimator of this density. We propose a cross-validation scheme for selecting an optimal bandwidth, which is proven to be a consistent procedure to estimate the density.

7.2.7. On the choice of weight functions for linear representations of persistence diagrams

Participant: Vincent Divol.

In collaboration with Wolfgang Polonik (UC Davis)

Persistence diagrams are efficient descriptors of the topology of a point cloud. As they do not naturally belong to a Hilbert space, standard statistical methods cannot be directly applied to them. Instead, feature maps (or representations) are commonly used for the analysis. A large class of feature maps, which we call linear, depends on some weight functions, the choice of which is a critical issue. An important criterion to choose a weight function is to ensure stability of the feature maps with respect to Wasserstein distances on diagrams. In [42], we improve known results on the stability of such maps, and extend it to general weight functions. We also address the choice of the weight function by considering an asymptotic setting; assume that X_n is an i.i.d. sample from a density on $[0,1]^d$. For the Cech and Rips filtrations, we characterize the weight functions for which the corresponding feature maps converge as n approaches infinity, and by doing so, we prove laws of large numbers for the total persistence of such diagrams. Both approaches lead to the same simple heuristic for tuning weight functions: if the data lies near a d-dimensional manifold, then a sensible choice of weight function is the persistence to the power α with $\alpha \geq d$.

7.2.8. Estimating the Reach of a Manifold

Participants: Frédéric Chazal, Bertrand Michel.

In collaboration with E. Aamari (CNRS Paris 7), J.Kim, A. Rinaldo and L. Wasserman (Carnegie Mellon University).

Various problems in manifold estimation make use of a quantity called the reach, denoted by τ_M, which is a measure of the regularity of the manifold. [32] is the first investigation into the problem of how to estimate the reach. First, 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 $\mathbb{X}_n$, $\tau(\mathbb{X}_n)$ is showed to achieve uniform expected loss bounds over a $\mathcal{C}^3$-like model. Finally, we obtain upper and lower bounds on the minimax rate for estimating the reach.

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

Participants: Frédéric Chazal, Bertrand Michel.

In collaboration with B. Fasy (Univ. Montana) and F. Lecci, A. Rinaldo and L. Wasserman (Carnegie Mellon University).

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), introduced by Chazal et al. (2011), and the kernel distance, introduced by Phillips et al. (2014), are smooth functions that provide useful topological information but are robust to noise and outliers. Chazal et al. (2015) derived concentration bounds for DTM. Building on these results, in [16], we derive limiting distributions and confidence sets, and we propose a method for choosing tuning parameters.

7.3. Topological approach for multimodal data processing

7.3.1. Barcode Embeddings for Metric Graphs

Participants: Steve Oudot, Yitchzak Solomon.

Stable topological invariants are a cornerstone of persistence theory and applied topology, but their discriminative properties are often poorly-understood. In [46] we study a rich homology-based invariant first defined by Dey, Shi, and Wang, which we think of as embedding a metric graph in the barcode space. We prove that this invariant is locally injective on the space of metric graphs and globally injective on a GH-dense subset. Moreover, we define a new topology on MGraphs, which we call the fibered topology, for which the barcode transform is injective on a generic (open and dense) subset.

7.3.2. Inverse Problems in Topological Persistence: a Survey

Participants: Steve Oudot, Yitchzak Solomon.

In [47] we review the literature on inverse problems in topological persistence theory. The first half of the survey is concerned with the question of surjectivity, i.e. the existence of right inverses, and the second half focuses on injectivity, i.e. left inverses. Throughout, we highlight the tools and theorems that underlie these advances, and direct the reader's attention to open problems, both theoretical and applied.

7.4. Experimental research and software development

7.4.1. Activity recognition from stride detection: a machine learning approach based on geometric patterns and trajectory reconstruction.

Participants: Bertrand Beaufils, Frédéric Chazal, Bertrand Michel.

In collaboration with M. Grelet (Sysnav).

In [23] algorithm for activity recognition is proposed using inertial sensors worn on the ankle. This innovative approach based on geometric patterns uses a stride detector that 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. A technique inspired by Zero Velocity Update is used on the stride detection to compute the trajectory of the device. It allows to compute relevant features for the activity recognition learning task. Compared to most algorithms in the literature, this method does not use fixed-size sliding window that could be too short to provide enough information or too long and leads to overlapping issue when the window covers two different activities.

7.4.2. Dynamics of silo deformation under granular discharge

Participant: Miroslav Kramár.

In collaboration with Claudia Colonnello.

In [17], we use Topological Data Analysis to study the post buckling behavior of laboratory scale cylindrical silos under gravity driven granular discharges. Thin walled silos buckle during the discharge if the initial height of the granular column is large enough. The deformation of the silo is reversible as long as the filling height does not exceed a critical value, Lc. Beyond this threshold the deformation becomes permanent and the silo often collapses. We study the dynamics of reversible and irreversible deformation processes, varying the initial filling height around Lc. We find that all reversible processes exhibit striking similarities and they alternate between regimes of slow and fast dynamics. The patterns that occur at the beginning of irreversible deformation processes are topologically very similar to those that arise during reversible processes. However, the dynamics of reversible and irreversible processes is significantly different. In particular, the evolution of irreversible processes is much faster. This allows us to make an early prediction of the collapse of the silo based solely on observations of the deformation patterns.

7.4.3. Characterizing Granular Networks Using Topological Metrics

Participant: Miroslav Kramár.

In collaboration with Joshua Dijksman (Duke Physics), Lenka Kovalcinova and Lou Kondic (NJIT), Jie Ren (Merck Research Lab), Robert Behringer (Duke), and Konstantin Mischaikow (Rutgers).

In [18], we carry out a direct comparison of experimental and numerical realizations of the exact same granular system as it undergoes shear jamming. We adjust the numerical methods used to optimally represent the experimental settings and outcomes up to microscopic contact force dynamics. Measures presented here range form microscopic, through mesoscopic to system-wide characteristics of the system. Topological properties of the mesoscopic force networks provide a key link between mi-cro and macro scales. We report two main findings: the number of particles in the packing that have at least two contacts is a good predictor for the mechanical state of the system, regardless of strain history and packing density. All measures explored in both experiments and numerics, including stress tensor derived measures and contact numbers depend in a universal manner on the fraction of non-rattler particles, fNR. The force network topology also tends to show this universality, yet the shape of the master curve depends much more on the details of the numerical simulations. In particular we show that adding force noise to the numerical data set can significantly alter the topological features in the data. We conclude that both fNR and topological metrics are useful measures to consider when quantifying the state of a granular system.

7.5. Miscellaneous

7.5.1. On Order Types of Random Point Sets

Participant: Marc Glisse.

In collaboration with Olivier Devillers and Xavier Goaoc (Inria team Gamble) and Philippe Duchon (LaBRI, Université de Bordeaux).

Let P be a set of n random points chosen uniformly in the unit square. In this paper [41], we examine the typical resolution of the order type of P. First, we show that with high probability, P can be rounded to the grid of step $\frac{1}{n^{3+\epsilon}}$ without changing its order type. Second, we study algorithms for determining the order type of a point set in terms of the number of coordinate bits they require to know. We give an algorithm that requires on average $4n\log_2 n + O(n)$ bits to determine the order type of P, and show that any algorithm requires at least $4n\log_2 n - O(n\log\log n)$ bits. Both results extend to more general models of random point sets.

GAIA Team

7. New Results

7.1. Regular (differential) chains

[17] provides new equivalence theorems for regular chains and regular differential chains, which are generalizations of Ritt's characteristic sets. These theorems focus on regularity properties of elements of residue class rings defined by these chains, which are revealed by resultant computations. New corollaries to these theorems have quite simple formulations.

[30] contains a description of the management of the parameters in the`Maple DifferentialAlgebra package` and, in particular, in the RosenfeldGroebner function.

7.2. Systems of integro-differential equations

[28], [29] present a proof of concept for symbolic and numeric methods dedicated to the parameter estimation problem for models formulated by means of nonlinear integro-differential equations. In particular, we address the computation of the model input-output equation and the numerical integration of integro-differential systems (the BLINEIDE library).

7.3. Certified non-conservative tests for the structural stability of discrete multidimensional systems

In collaboration with Fabrice Rouillier (Inria Paris, Ouragan), in [18], we propose a new approach for testing the stability of nD systems. We first show that the standard characterization of the structural stability of a multivariate rational transfer function (namely, the denominator of the transfer function does not have solutions in the unit polydisc of $\mathbb{C}^n$) is equivalent to the fact that a certain system of polynomials does not have real solutions. We then use state-of-the-art computer algebra algorithms to check this last condition, and thus the structural stability of multidimensional systems. Our results have been implemented in a `Maple` prototype.

7.4. Using symbolic computation to solve algebraic Riccati equations arising in invariant filtering

In this joint work with Axel Barrau from Safran Tech [23], we propose a new step in the development of invariant observers. In the past, this theory led to impressive simplifications of the error equations encountered in estimation problems, especially those related to navigation. This was used to reduce computation load or derive new theoretical properties. Here, we leverage this advantage to obtain closed-form solutions of the underlying algebraic Riccati equations through advanced symbolic computation methods.

7.5. Parametric sub-optimal H_∞ controllers for an optro-mechanical system

In collaboration with *Safran Electronics & Defense*, in [15], we studied the robust stabilization of the line of sight of a stabilized mirror system. This system can be modeled by a single-input single- output time-delay system. Due to large model uncertainties, non-parametric methods are usually too conservative. Hence, we consider here unfixed model parameters. Using an additive decomposition, we show how to compute parametric H_∞ controllers of the time-delay model. Such a symbolic approach is interesting in the context of adaptive control and is illustrated throughout a simulation with an ideal parameter estimator.

7.6. A symbolic approach for signal demodulation and application to gearbox vibration analysis

This work is made in collaboration with Axel Barrau and Elisa Hubert (Safran Tech), and Roudy Dagher (Research Engineer, Inria Chile). The problem under study, which reduces to a certain signal factorization problem, was shown by Barrau et. al. to be equivalent to a Frobenius norm minimization problem. Starting from this optimization problem, we investigate the use of computer algebra methods to compute explicit solutions for the original problem. Along the way, we exhibit interesting algebraic and geometric properties of the underlying polynomial system. A paper is currently in development to summarize these results.

7.7. Curve analysis for the stability of time-delay systems

This work aims to design a new symbolic-numerical Puiseux-free approach for the study of the stability of differential time-delay systems. The idea behind is to replace the costly computations of Puiseux developpements around the *critical pairs* of the characteristic function by the numerical analysis of the branches of a well chosen 3D curve. The preliminary results show that this approach is easier to implement and turns out to be more efficient in practice. This ongoing work will be the subject of a future publication.

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 [6] 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 [49], 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 showed that our reliable numerical approach can handle classes of examples that are not reachable by symbolic methods.

7.1.2. Workspace, Joint space and Singularities of a family of Delta-Like Robots

Our paper [7] presents the workspace, the joint space and the singularities of a family of delta-like parallel robots by using algebraic tools. The different functions of the SIROPA library are introduced and used to estimate the complexity representing the singularities in the workspace and the joint space. A Groebner based elimination is used to compute the singularities of the manipulator and a Cylindrical Algebraic Decomposition algorithm is used to study the workspace and the joint space. From these algebraic objects, we propose some certified three-dimensional plotting tools describing the shape of the workspace and of the joint space which will help engineers or researchers to decide the most suited configuration of the manipulator they should use for a given task. Also, the different parameters associated with the complexity of the serial and parallel singularities are tabulated, which further enhance the selection of the different configurations of the manipulator by comparing the complexity of the singularity equations.

In collaboration with Ranjan Jha, Damien Chablat, Luc Baron and Fabrice Rouillier.

7.2. Non-Euclidean Computational Geometry

Participants: Vincent Despré, Iordan Iordanov, Monique Teillaud.

7.2.1. Delaunay Triangulations of Symmetric Hyperbolic Surfaces

We have worked on extending our previous results on the computation of Delaunay triangulations of the Bolza surface [50] (see also the section New Software above), which is the most symmetric surface of genus 2. Elaborating further on previous work [26], we are now considering symmetric hyperbolic surfaces of higher genus, for which we study mathematical properties [14] that allow us to propose algorithms [13].

In collaboration with Gert Vegter and Matthijs Ebbens (University of Groningen).

7.3. Probabilistic Analysis of Geometric Data Structures and Algorithms

Participants: Olivier Devillers, Charles Duménil, Fernand Kuiebove Pefireko.

7.3.1. Stretch Factor in a Planar Poisson-Delaunay Triangulation with a Large Intensity

Let $X := X_n \cup \{(0,0),(1,0)\}$, where X_n is a planar Poisson point process of intensity n. Our paper [4] provides a first non-trivial lower bound for the expected length of the shortest path between $(0,0)$ and $(1,0)$ in the Delaunay triangulation associated with X when the intensity of X_n goes to infinity. Simulations indicate that the correct value is about 1.04. We also prove that the expected length of the so-called upper path converges to $\frac{35}{3\pi^2}$, giving an upper bound for the expected length of the smallest path.

In collaboration with Nicolas Chenavier (Université du Littoral Côte d'Opale).

7.3.2. Delaunay triangulation of a Poisson Point Process on a Surface

The complexity of the Delaunay triangulation of n points distributed on a surface ranges from linear to quadratic. We proved that when the points are evenly distributed on a smooth compact generic surface the expected size of the Delaunay triangulation is can be controlled. If the point set is a good sample of a smooth compact generic surface [22] the complexity is controlled. Namely, good sample means that a sphere of size ϵ centered on the surface contains between 1 and η points. Under this hypothesis, the complexity of the Delaunay triangulation is $O\left(\frac{\eta^2}{\epsilon^2}\log\frac{1}{\epsilon}\right)$. We proved that when the points are evenly distributed on a smooth compact generic surface they form a good sample with high probability for relevant values of ϵ and η. We can deduce [15] that the expected size of the Delaunay triangulation of n random points of a surface is $O(n\log^2 n)$.

7.3.3. On Order Types of Random Point Sets

Let P be a set of n random points chosen uniformly in the unit square. In our paper [19], we examine the typical resolution of the order type of P. First, we showed that with high probability, P can be rounded to the grid of step $\frac{1}{n^{3+\epsilon}}$ without changing its order type. Second, we studied algorithms for determining the order type of a point set in terms of the the number of coordinate bits they require to know. We gave an algorithm that requires on average $4n\log_2 n + O(n)$ bits to determine the order type of P, and showed that any algorithm requires at least $4n\log_2 n - O(n\log\log n)$ bits. Both results extend to more general models of random point sets.

In collaboration with Philippe Duchon (LABRI) and Marc Glisse (project team DATASHAPE *).*

7.4. Classical Computational Geometry and Graph Drawing

Participants: Vincent Despré, Olivier Devillers, Sylvain Lazard.

7.4.1. Delaunay Triangulations of Points on Circles

Delaunay triangulations of a point set in the Euclidean plane are ubiquitous in a number of computational sciences, including computational geometry. Delaunay triangulations are not well defined as soon as 4 or more points are concyclic but since it is not a generic situation, this difficulty is usually handled by using a (symbolic or explicit) perturbation. As an alternative, we proposed to define a canonical triangulation for a set of concyclic points by using a max-min angle characterization of Delaunay triangulations. This point of view leads to a well defined and unique triangulation as long as there are no symmetric quadruples of points. This unique triangulation can be computed in quasi-linear time by a very simple algorithm [18].

In collaboration with Hugo Parlier and Jean-Marc Schlenker (University of Luxembourg).

7.4.2. Improved Routing on the Delaunay Triangulation

A geometric graph $G=(P,E)$ is a set of points in the plane and edges between pairs of points, where the weight of each edge is equal to the Euclidean distance between the corresponding points. In k-local routing we find a path through G from a source vertex s to a destination vertex t, using only knowledge of the present location, the locations of s and t, and the k-neighbourhood of the current vertex. We presented [11] an algorithm for 1-local routing on the Delaunay triangulation, and show that it finds a path between a source vertex s and a target vertex t that is not longer than $3.56|st|$, improving the previous bound of 5.9.

In collaboration with Nicolas Bonichon (Labri), Prosenjit Bose, Jean-Lou De Carufel, Michiel Smid and Daryl Hill (Carleton University)

7.4.3. Limits of Order Types

We completed an extended version of a work published at SoCG 2015, in which we apply ideas from the theory of limits of dense combinatorial structures to study order types, which are combinatorial encodings of finite point sets. Using flag algebras we obtain new numerical results on the Erdös problem of finding the minimal density of 5-or 6-tuples in convex position in an arbitrary point set, and also an inequality expressing the difficulty of sampling order types uniformly. Next we establish results on the analytic representation of limits of order types by planar measures. Our main result is a rigidity theorem: we show that if sampling two measures induce the same probability distribution on order types, then these measures are projectively equivalent provided the support of at least one of them has non-empty interior. We also show that some condition on the Hausdorff dimension of the support is necessary to obtain projective rigidity and we construct limits of order types that cannot be represented by a planar measure. Returning to combinatorial geometry we relate the regularity of this analytic representation to the aforementioned problem of Erdös on the density of k-tuples in convex position, for large k [20].

In collaboration with Alfredo Hubard (Laboratoire d'Informatique Gaspard-Monge) Rémi De Joannis de Verclos (Radboud university, Nijmegen) Jean-Sébastien Sereni (CNRS) Jan Volec (Department of Mathematics and Computer Science, Emory University)

7.4.4. 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 case, the size of $\mathcal{Q}$ is $O(n^{15})$, but we conjecture a good complexity of $O(n\sqrt{n})$ in practice on non-pathological data [12].

In collaboration with William J. Lenhart (Williams College, USA).

7.4.5. On the Edge-length Ratio of Outerplanar Graphs

We show that any outerplanar graph admits a planar straight-line drawing such that the length ratio of the longest to the shortest edges is strictly less than 2. This result is tight in the sense that for any $\epsilon > 0$ there are outerplanar graphs that cannot be drawn with an edge-length ratio smaller than $2 - \epsilon$. We also show that this ratio cannot be bounded if the embeddings of the outerplanar graphs are given [9].

In collaboration with William J. Lenhart (Williams College, USA) and Giuseppe Liotta (Università di Perugia, Italy).

GRACE Project-Team

6. New Results

6.1. Fast transforms over fields of characteristic 2

Participant: Nicholas Coxon.

With the aim of reaching fast, linear time, algorithms for encoding multiplicity codes, which have good local properties, N. Coxon had to develop subalgorithms for dealing with the Hermite interpolation [13], which in turn relies on computer algebra for fast transforms over fields of characteritic two [14]. Locally decodable codes are used for private information retrieval, where a database can be privately queried by a user, in such a way that the user does not reveal his query. Using codes with locality for private information retrieval, the database is first encoded, then queried using the local property of the code. Since the databases in question can be large, only linear time algorithms can be used. Our results achieve linear-time complexity, and even with a non agressively optimized implementation, can encode as much as 10^9 bits in thirty seconds on a laptop.

6.2. Private information retrieval

Participants: Daniel Augot, Nicholas Coxon, Julien Lavauzelle, Françoise Levy-Dit-Vehel.

J. Lavauzelle continued his study on private information retrieval (PIR) protocols. First, he completed the construction of PIR protocols from transversal designs [8], initiated in 2017. Compared to existing protocols, the main benefit of the construction is to feature an optimal computation complexity for the servers. Sublinear communication complexity and negligeable storage overhead can also be achieved for some particular instances.

Second, in a joint work with R. Tajeddine, R. Freij-Hollanti and C. Hollanti from the University of Aalto (Finland), J. Lavauzelle considered the setting in which the database is encoded with an optimal regenerating code [16]. Quantitatively, their construction of PIR protocols improves upon a recent work of Dorkson and Ng, for every non-trivial set of parameters.

6.3. Locally correctable codes

Participant: Julien Lavauzelle.

In 2013, Guo, Kopparty and Sudan built a new family of locally correctable codes from lifting, achieving an arbitrarily high information rate for sublinear locality. J. Lavauzelle proposed an analogue of this construction in projective spaces [7]. The parameters of this construction are similar to the original work of Guo *et al.* Intertwined relations between the two families of codes were proven thanks to a careful analysis of their monomial bases. The practicality of the construction was also established through an implementation and a study of information sets and automorphisms of the code.

6.4. Cryptanalysis in code based cryptography

Participant: Alain Couvreur.

Following NIST call for post quantum cryptography, A. Couvreur and E. Barelli designed a key recovery attack against a McEliece–like encryption scheme called DAGS [9].

In addition, in collaboration with Matthieu Lequesne and Jean-Pierre Tillich (Inria Paris, SECRET team), A. Couvreur designed an attack against another proposal called RLCE (Random Linear Code Encryption) [12].

6.5. Commutative isogeny-based cryptography

Participants: Luca de Feo, Benjamin Smith.

Despite the many advances in post-quantum cryptography in recent years, efficient drop-in replacements for the classic Diffie–Hellman key exchange algorithm have proven elusive. L. De Feo, J. Kieffer, and B. Smith laid the algorithmic groundwork for *commutative isogeny-based key exchange* in [10]; this work became the basis of the exciting new CSIDH proposal [19].

6.6. Factoring oracles

Participants: François Morain, Benjamin Smith.

Integer factoring is an old topic, and the situation is as follows: in the classical world, we think integer factoring is hard and the algorithms we have are quite powerful though of subexponential complexity and factoring numbers with several hundred bits; whereas in the quantum world, it is assumed to be easy (i.e., there exists a quantum polynomial time algorithm) but never experienced and the record is something like a few bits. F. Morain, helped by B. Smith and G. Renault (ANSSI) studied the theoretical problem of factoring integers given access to classical oracles, like the Euler totient function. They were able to give some interesting classes of numbers that could tackled, see [17].

LFANT Project-Team

6. New Results

6.1. Cryptographic Protocols

Participants: Guilhem Castagnos, Ida Tucker.

In [24], G. Castagnos, F. Laguillaumie and I. Tucker revisit a recent cryptographic primitive called *Functional encryption for inner products* (FE4IP).

Functional encryption (FE) is an advanced cryptographic primitive which allows, for a single encrypted message, to finely control how much information on the encrypted data each receiver can recover. To this end many functional secret keys are derived from a master secret key. Each functional secret key allows, for a ciphertext encrypted under the associated public key, to recover a specific function of the underlying plaintext.

Since constructions for general FE that appear in the past five years are far from practical, the problem arose of building efficient FE schemes for restricted classes of functions; and in particular for linear functions, (i.e. the inner product functionality). Such constructions yield many practical applications, while developing our understanding of FE.

Though such schemes had already been conceived in the past three years (Abdalla *et al.* 2015, Agrawal *et al.* 2016), they all suffered of practical drawbacks. Namely the computation of inner products modulo a prime are restricted, in that they require that the resulting inner product be small for decryption to be efficient. The only existing scheme that overcame this constraint suffered of poor efficiency due in part to very large ciphertexts. This work overcomes these limitations and we build the first FE schemes for inner products modulo a prime that are both efficient and recover the result whatever its size.

To this end, Castagnos *et al.* introduce two new cryptographic assumptions. These are variants of the assumptions used for the Castagnos-Laguillaumie encryption of 2015. This supposes the existence of a cyclic group G where the decision Diffie-Hellman assumption holds together with a subgroup F of G where the discrete logarithm problem is easy. This setting allows to encode information in the exponent of the subgroup F, which can be efficiently recovered whatever its size.

From these assumptions Castagnos *et al.* construct generic, linearly homomorphic encryption schemes over a field of prime order which are semantically secure under chosen plaintext attacks. They then use the homomorphic properties of the above schemes to construct generic inner product FE schemes over the integers and over fields of prime order. They thereby provide constructions for inner product FE modulo a prime p that do not restrict the size of the inputs or of the resulting inner product, which are the most efficient such schemes to date.

This paper was presented at the ASIACRYPT Conference 2018, and is part of the ALAMBIC project.

6.2. Computation of Euclidean minima in totally definite quaternion fields

Participant: Jean-Paul Cerri.

In collaboration with Pierre Lezowski, Jean-Paul Cerri has studied norm-Euclidean properties of totally definite quaternion fields over number fields. Building on their previous work about number fields, they have proved that the Euclidean minimum and the inhomogeneous minimum of orders in such quaternion fields are always equal. Besides, they are rational under the hypothesis that the base number field is not quadratic. This single remaning open case corresponds to the similar open case remaining for real number fields.

They also have extended Cerri's algorithm for the computation of the upper part of the norm-Euclidean spectrum of a number field to this noncommutative context. This algorithm has allowed to compute the exact value of the norm-Euclidean minimum of orders in totally definite quaternion fields over a quadratic number field. This has provided the first known values of this minimum when the base number field has degree strictly greater than 1.

Consequently, both theoretical and practical milestones set in the previous quadrennial report were reached. These results are presented in [19], due to appear in *International Journal of Number Theory*.

6.3. Can you hear the homology of 3-dimensional drums?

Participant: Aurel Page.

In [16], A. Bartel and A. Page describe all possible actions of groups of automorphisms on the homology of 3-manifolds, and prove that for every prime p, there are 3-dimensional drums that sound the same but have different p-torsion in their homology. This completes previous work [42] by proving that the behaviour observed by computer experimentation was indeed a general phenomenon.

More precisely: if M is a manifold with an action of a group G, then the homology group $H_1(M,\mathbb{Q})$ is naturally a $\mathbb{Q}[G]$-module, where $\mathbb{Q}[G]$ denotes the rational group ring. Bartel and Page prove that for every finite group G, and for every $\mathbb{Q}[G]$-module V, there exists a closed hyperbolic 3-manifold M with a free G-action such that the $\mathbb{Q}[G]$-module $H_1(M,\mathbb{Q})$ is isomorphic to V. They give an application to spectral geometry: for every finite set P of prime numbers, there exist hyperbolic 3-manifolds N and N' that are strongly isospectral such that for all $p\in P$, the p-power torsion subgroups of $H_1(N,\mathbb{Z})$ and of $H_1(N',\mathbb{Z})$ have different orders. They also show that, in a certain precise sense, the rational homology of oriented Riemannian 3-manifolds with a G-action "knows" nothing about the fixed point structure under G, in contrast to the 2-dimensional case. The main geometric techniques are Dehn surgery and, for the spectral application, the Cheeger-Müller formula, but they also make use of tools from different branches of algebra, most notably of regulator constants, a representation theoretic tool that was originally developed in the context of elliptic curves.

6.4. Error-correcting codes based on non-commutative algebras

Participant: Aurel Page.

In [36], C. Maire and A. Page revisit a construction due to Lenstra and Guruswami by generalising them to unit groups of division algebras.

Lenstra and Guruswami described number field analogues of the algebraic geometry codes of Goppa. Recently, Maire and Oggier generalised these constructions to other arithmetic groups: unit groups in number fields and orders in division algebras; they suggested to use unit groups in quaternion algebras but could not completely analyse the resulting codes. Maire and Page prove that the noncommutative unit group construction yields asymptotically good families of codes for division algebras of any degree, and estimate the smallest possible size of the alphabet in terms of the degree of the algebra.

6.5. Towards practical key exchange from ordinary isogeny graphs

Participant: Jean Kieffer.

In [25], L. De Feo, J. Kieffer and B. Smith revisit the ordinary isogeny-graph based cryptosystems of Couveignes and Rostovtsev–Stolbunov, long dismissed as impractical.

De Feo, Kieffer and Smith give algorithmic improvements that accelerate key exchange in this framework, and explore the problem of generating suitable system parameters for contemporary pre-and post-quantum security that take advantage of these new algorithms. They prove the session-key security of this key exchange in the Canetti-Krawczyk model, and the IND-CPA security of the related public-key encryption scheme, under reasonable assumptions on the hardness of computing isogeny walks. This system admits efficient key-validation techniques that yield CCA-secure encryption, thus providing an important step towards efficient post-quantum non-interactive key exchange (NIKE).

6.6. Optimal addition sequences for theta functions

Participants: Andreas Enge, Fredrik Johansson.

In [20], A. Enge, F. Johansson and their coauthor W. Hart consider the problem of numerically evaluating one-dimensional θ-functions and the elliptic η-function. They construct short addition sequences reaching an optimal number of $N + o(N)$ multiplications for evaluating the function as a sparse series with N terms. The proof relies on the representability of specific quadratic progressions of integers as sums of smaller numbers of the same kind. For example, they show that every generalised pentagonal number $c > 5$ can be written as $c = 2a + b$, where a, b are smaller generalised pentagonal numbers. They then give a baby-step giant-step algorithm that breaks through the theoretical barrier achievable with addition sequences, and which uses only $O(N/(logN)^r)$ multiplications for any $r > 0$. These theoretical improvements also lead to an interesting speed-up in practice, and they have been integrated into the CM and the ARB software.

6.7. Reed–Solomon-Gabidulin Codes

Participant: Xavier Caruso.

In [31], X. Caruso and A. Durand define a new family of linear codes which is a common generalization of Reed–Solomon codes on the one hand and Gabidulin codes on the other hand. Their construction works over an arbitrary field (not necessarily finite) equipped with an automorphism of finite order and a twisted derivation whose subfield of constants is sufficiently large. This setting allows for example the base field to be $\mathbb{F}_q(t)$ equipped with its natural derivation and then provides a new large family of interesting codes. Caruso and Durand then compute the minimal distance of their codes and design an efficient algorithm for decoding up to the half of the minimal distance.

6.8. Computing Stieltjes constants using complex integration

Participant: Fredrik Johansson.

In [32], F. Johansson and I. Blagouchine devise an efficient algorithm to compute the generalized Stieltjes constants $\gamma_n(a)$ to arbitrary precision with rigorous error bounds, for the first time achieving this with low complexity with respect to the order n. The algorithm consists of locating an approximate steepest descent contour and then evaluating the integral numerically in ball arithmetic using the Petras algorithm with a Taylor expansion for bounds near the saddle point. An implementation is provided in the Arb library.

6.9. Numerical Evaluation of Elliptic Functions, Elliptic Integrals and Modular Forms

Participant: Fredrik Johansson.

In [33], F. Johansson describes algorithms to compute elliptic functions and their relatives (Jacobi theta functions, modular forms, elliptic integrals, and the arithmetic-geometric mean) numerically to arbitrary precision with rigorous error bounds for arbitrary complex variables. Implementations in ball arithmetic are available in the Arb library. This overview article discusses the standard algorithms from a concrete implementation point of view, and also presents some improvements.

6.10. Numerical integration in arbitrary-precision ball arithmetic

Participant: Fredrik Johansson.

In [26], F. Johansson describes an implementation of arbitrary-precision numerical integration with rigorous error bounds in the Arb library. Rapid convergence is ensured for piecewise complex analytic integrals by use of the Petras algorithm, which combines adaptive bisection with adaptive Gaussian quadrature where error bounds are determined via complex magnitudes without evaluating derivatives. The code is general, easy to use, and efficient, often outperforming existing non-rigorous software.

6.11. Fast and rigorous arbitrary-precision computation of Gauss-Legendre quadrature nodes and weights

Participant: Fredrik Johansson.

In [26], F. Johansson and M. Mezzarobba describe a strategy for rigorous arbitrary-precision evaluation of Legendre polynomials on the unit interval and its application in the generation of Gauss-Legendre quadrature rules. The focus is on making the evaluation practical for a wide range of realistic parameters, corresponding to the requirements of numerical integration to an accuracy of about 100 to 100 000 bits. The algorithm combines the summation by rectangular splitting of several types of expansions in terms of hypergeometric series with a fixed-point implementation of Bonnet's three-term recurrence relation. Rigorous enclosures of the Gauss-Legendre nodes and weights are then computed using the interval Newton method. The work provides rigorous error bounds for all steps of the algorithm. The approach is validated by an implementation in the Arb library, which achieves order-of-magnitude speedups over previous code for computing Gauss-Legendre rules with simultaneous high degree and precision.

6.12. On a two-valued sequence and related continued fractions in power series fields

Participant: Bill Allombert.

In [15], Bill Allombert with Nicolas Brisebarre and Alain Lasjaunias describe a noteworthy transcendental continued fraction in the field of power series over $\mathbb{Q}$, having irrationality measure equal to 3. This article has been published in The Ramanujan Journal.

6.13. Moduli space

Participant: Nicolas Mascot.

The article [22] by Nicolas Mascot, on the Certification of modular Galois representations has been published in Mathematics of Computation.

6.14. Modular forms

Participants: Karim Belabas, Henri Cohen, Bill Allombert.

In [18], K. Belabas and H. Cohen give theoretical and practical information on the Pari/GP modular forms package, using the formalism of trace formulas. This huge package (about 70 exported public functions) handles standard operations on classical modular forms in $M_k(\Gamma_0(N), \chi)$, also in weight 1 and non-integral weight (which are not cohomological, hence not directly handled by trace formulas). It is the first publicly available package which can compute Fourier expansions at any cusps, evaluate modular forms near the real axis, evaluate L-functions of non-eigenforms, and compute general Petersson scalar products.

In [39], H. Cohen explained how to compute Fourier expansions at all cusps of any modular form of integral or half-integral weight.

A complementary package using modular symbols is used in [17] by Karim Belabas, Dominique Bernardi and Bernadette Perrin-Riou to compute Manin's constant and the modular degree of elliptic curves defined over $\mathbb{Q}$.

6.15. L-functions

Participant: Henri Cohen.

In [29], H. Cohen gives an overview of Computational Number Theory in Relation with L-Functions, both in the local case (counting points on varieties over finite fields, involving in particular a detailed study of Gauss and Jacobi sums), and in the global case (for instance Dirichlet L-functions, involving in particular the study of inverse Mellin transforms). He also gives a number of little-known but very useful numerical methods, usually but not always related to the computation of L-functions.

6.16. Number fields

Participant: Henri Cohen.

In https://hal.inria.fr/hal-01379473/, H. Cohen and F. Thorne give explicit formulas for the Dirichlet series generating function of D_ℓ-extensions of odd prime degree ℓ with given quadratic resolvent.

OURAGAN Team

7. New Results

7.1. On $SL(3,\mathbb{C})$-representations of the Whitehead link group

In [9], we describe a family of representations in $SL(3,\mathbb{C})$ of the fundamental group π of the Whitehead link complement. These representations are obtained by considering pairs of regular order three elements in $SL(3,\mathbb{C})$ and can be seen as factorising through a quotient of π defined by a certain exceptional Dehn surgery on the Whitehead link. Our main result is that these representations form an algebraic component of the $SL(3,\mathbb{C})$-character variety of π.

7.2. A simplified approach to rigorous degree 2 elimination in discrete logarithm algorithms

In [10], we revisit the ZigZag strategy of Granger, Kleinjung and Zumbrägel. In particular, we provide a new algorithm and proof for the so-called degree 2 elimination step. This allows us to provide a stronger theorem concerning discrete logarithm computations in small characteristic fields $F_{q^{k_0k}}$ with k close to q and k_0 a small integer. As in the aforementioned paper, we rely on the existence of two polynomials h_0 and h_1 of degree 2 providing a convenient representation of the finite field $F_{q^{k_0k}}$.

7.3. Computing Chebyshev knot diagrams

A Chebyshev curve $\mathcal{C}(a,b,c,\phi)$ has a parametrization of the form $x(t)=T_a(t);\quad y(t)=T_b(t);$ $z(t)=T_c(t+\phi)$, where a,b,c are integers, $T_n(t)$ is the Chebyshev polynomial of degree n and $\phi\in\mathbb{R}$. When $\mathcal{C}(a,b,c,\phi)$ is nonsingular, it defines a polynomial knot. In [12], we determine all possible knot diagrams when ϕ varies. Let a,b,c be integers, a is odd, $(a,b)=1$, we show that one can list all possible knots $\mathcal{C}(a,b,c,\phi)$ in $O(n^2)$ bit operations, with $n=abc$.

7.4. Programmable projective measurement with linear optics

In [8] present a scheme for a universal device which can be programmed by quantum states to perform a chosen projective measurement, and its implementation in linear optics. In particular, our scheme takes a single input system (the input register), and M-1 systems all in a state psi (the program registers), whose role is to encode the measurement direction, and approximates the projective measurement with respect to the state psi on the input system. Importantly the scheme is entirely independent of the measurement basis choice psi. This is done optimally in M, if we demand the input state psi always returns the appropriate outcome, and limits to the ideal projective measurement with M. The size of the linear optical circuit we propose scales as M log M , and requires O(M log M) classical side processing. Our scheme can be viewed as an extension of the swap test to the instance where one state is supplied many times.

7.5. Updating key size estimations for pairings

Recent progress on NFS imposed a new estimation of the security of pairings. In [6], we study the best attacks against some of the most popular pairings. It allows us to propose new pairing-friendly curves of 128 bits and 192 bits of security.

7.6. How to Securely Compute with Noisy Leakage in Quasilinear Complexity

Since their introduction in the late 90's, side-channel attacks have been considered as a major threat against cryptographic implementations. This threat has raised the need for formal leakage models in which the security of implementations can be proved. At Eurocrypt 2013, Prouff and Rivain introduced the noisy leakage model which has been argued to soundly capture the physical reality of power and electromagnetic leakages. In their work, they also provide the first formal security proof for a masking scheme in the noisy leakage model. However their work has two important limitations: (i) the security proof relies on the existence of a leak-free component, (ii) the tolerated amount of information in the leakage (aka leakage rate) is of $O(1/n)$ where n is the number of shares in the underlying masking scheme. The first limitation was nicely tackled by Duc, Dziembowski and Faust one year later (Eurocrypt 2014). Their main contribution was to show a security reduction from the noisy leakage model to the conceptually simpler random- probing model. They were then able to prove the security of the well-known Ishai-Sahai-Wagner scheme (Crypto 2003) in the noisy leakage model. The second limitation was addressed last year in a paper by Andrychowicz, Dziembowski and Faust (Eurocrypt 2016). The proposed construction achieves security in the strong adaptive probing model with a leakage rate of $O(1/logn)$ at the cost of a $O(n^2logn)$ complexity. we argue that their result can be translated into the noisy leakage model with a leakage rate of $O(1)$ by using secret sharing based on algebraic geometric codes. They further argue that the efficiency of their construction can be improved by a linear factor using packed secret sharing but no details are provided.

In [14], we show how to compute in the presence of noisy leakage with a leakage rate up to $\widetilde{O}(1)$ in complexity $\widetilde{O}(n)$. They use a polynomial encoding allowing quasilinear multiplication based on the fast Number Theoretic Transform (NTT). They first show that the scheme is secure in the random-probing model with leakage rate $O(1/logn)$. Using the reduction by Duc et al. this result can be translated in the noisy leakage model with a $O(1/|F|^2logn)$ leakage rate. However, as in the work of Andrychowicz et al. , our construction also requires $|F| = O(n)$. In order to bypass this issue, we refine the granularity of our computation by considering the noisy leakage model on logical instructions that work on constant-size machine words. we provide a generic security reduction from the noisy leakage model at the logical-instruction level to the random-probing model at the arithmetic level. This reduction allows to prove the security of the construction in the noisy leakage model with leakage rate $\widetilde{O}(1)$).

7.7. A New Public-Key Cryptosystem via Mersenne Numbers

In [13], we propose a new public-key cryptosystem whose security is based on the computational intractability of the following problem: Given a Mersenne number $p = 2^n - 1$ where n is a prime, a positive integer h , and two n -bit integers T,R , find two n -bit integers F,G each of Hamming weight at most h such that T = F ·R+G modulo p , under the promise that they exist.

7.8. Workspace, Joint space and Singularities of a family of Delta-Like Robot

In [11], we describe the workspace, the joint space and the singularities of a family of delta-like parallel robots by using algebraic tools. The different func- tions of SIROPA library are introduced, which is used to induce an estimation about the complexity in representing the singularities in the workspace and the joint space. A Gröbner based elimination is used to compute the singularities of the manipulator and a Cylindrical Algebraic Decomposition algorithm is used to study the workspace and the joint space. From these algebraic objects, they propose some certified three-dimensional plotting describing the shape of works- pace and of the joint space which will help the engineers or researchers to decide the most suited configuration of the manipulator they should use for a given task. Also, the different parameters associated with the complexity of the serial and parallel singularities are tabulated, which further enhance the selection of the different configuration of the manipulator by comparing the complexity of the singularity equations.

7.9. Certified Non-conservative Tests for the Structural Stability of Discrete Multidimensional Systems

In [7], we present new computer algebra based methods for testing the structural stability of n-D discrete linear systems (with $n \geq 2$). More precisely, they show that the standard characterization of the structural stability of a multivariate rational transfer function (namely, the denominator of the transfer function does not have solutions in the unit polydisc of $\mathbb{C}^n$) is equivalent to the fact that a certain system of polynomials does not have real solutions. We then use state-of-the-art computer algebra algorithms to check this last condition, and thus the structural stability of multidimensional systems.

POLSYS Project-Team

6. New Results

6.1. Fundamental algorithms and structured polynomial systems

6.1.1. Towards Mixed Gröbner Basis Algorithms: the Multihomogeneous and Sparse Case

One of the biggest open problems in computational algebra is the design of efficient algorithms for Gröbner basis computations that take into account the sparsity of the input polynomials. We can perform such computations in the case of unmixed polynomial systems, that is systems with polynomials having the same support, using the approach of Faugère, Spaenlehauer, and Svartz [ISSAC'14]. In [15] we present two algorithms for sparse Gröbner bases computations for mixed systems. The first one computes with mixed sparse systems and exploits the supports of the polynomials. Under regularity assumptions, it performs no reductions to zero. For mixed, square, and 0-dimensional multihomogeneous polynomial systems, we present a dedicated, and potentially more efficient, algorithm that exploits different algebraic properties that performs no reduction to zero. We give an explicit bound for the maximal degree appearing in the computations.

6.1.2. Bilinear Systems with Two Supports: Koszul Resultant Matrices, Eigenvalues, and Eigenvectors

A fundamental problem in computational algebraic geometry is the computation of the resultant. A central question is when and how to compute it as the determinant of a matrix whose elements are the coefficients of the input polynomials up-to sign. This problem is well understood for unmixed multihomogeneous systems, that is for systems consisting of multihomogeneous polynomials with the same support. However, little is known for mixed systems, that is for systems consisting of polynomials with different supports. In [14] we consider the computation of the multihomogeneous resultant of bilinear systems involving two different supports. We present a constructive approach that expresses the resultant as the exact determinant of a *Koszul resultant matrix*, that is a matrix constructed from maps in the Koszul complex. We exploit the resultant matrix to propose an algorithm to solve such systems. In the process we extend the classical eigenvalues and eigenvectors criterion to a more general setting. Our extension of the eigenvalues criterion applies to a general class of matrices, including the Sylvester-type and the Koszul-type ones.

6.1.3. A Polynomial-Division-Based Algorithm for Computing Linear Recurrence Relations

Sparse polynomial interpolation, sparse linear system solving or modular rational reconstruction are fundamental problems in Computer Algebra. They come down to computing linear recurrence relations of a sequence with the Berlekamp–Massey algorithm. Likewise, sparse multivariate polynomial interpolation and multidimensional cyclic code decoding require guessing linear recurrence relations of a multivariate sequence.

Several algorithms solve this problem. The so-called Berlekamp–Massey–Sakata algorithm (1988) uses polynomial additions and shifts by a monomial. The SCALAR-FGLM algorithm (2015) relies on linear algebra operations on a multi-Hankel matrix, a multivariate generalization of a Hankel matrix. The Artinian Gorenstein border basis algorithm (2017) uses a Gram-Schmidt process.

In [16], we propose a new algorithm for computing the Gröbner basis of the ideal of relations of a sequence based solely on multivariate polynomial arithmetic. This algorithm allows us to both revisit the Berlekamp–Massey–Sakata algorithm through the use of polynomial divisions and to completely revise the SCALAR-FGLM algorithm without linear algebra operations.

A key observation in the design of this algorithm is to work on the mirror of the truncated generating series allowing us to use polynomial arithmetic modulo a monomial ideal. It appears to have some similarities with Padé approximants of this mirror polynomial.

Finally, we give a partial solution to the transformation of this algorithm into an adaptive one.

As an addition from the paper published at the ISSAC conferance, in [24], we give an adaptive variant of this algorithm taking into account the shape of the final Gröbner basis gradually as it is discovered. The main advantage of this algorithm is that its complexity in terms of operations and sequence queries only depends on the output Gröbner basis.

All these algorithms have been implemented in MAPLE and we report on our comparisons.

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

The BERLEKAMP–MASSEY–SAKATA algorithm and the SCALAR-FGLM algorithm both compute the ideal of relations of a multidimensional linear recurrent sequence.

Whenever quering a single sequence element is prohibitive, the bottleneck of these algorithms becomes the computation of all the needed sequence terms. As such, having adaptive variants of these algorithms, reducing the number of sequence queries, becomes mandatory.

A native adaptive variant of the SCALAR-FGLM algorithm was presented by its authors, the so-called ADAPTIVE SCALAR-FGLM algorithm.

In [25], our first contribution is to make the BERLEKAMP–MASSEY–SAKATA algorithm more efficient by making it adaptive to avoid some useless relation testings. This variant allows us to divide by four in dimension 2 and by seven in dimension 3 the number of basic operations performed on some sequence family.

Then, we compare the two adaptive algorithms. We show that their behaviors differ in a way that it is not possible to tweak one of the algorithms in order to mimic exactly the behavior of the other. We detail precisely the differences and the similarities of both algorithms and conclude that in general the ADAPTIVE SCALAR-FGLM algorithm needs fewer queries and performs fewer basic operations than the ADAPTIVE BERLEKAMP–MASSEY–SAKATA algorithm.

We also show that these variants are always more efficient than the original algorithms.

6.1.5. Bit complexity for multi-homogeneous polynomial system solving Application to polynomial minimization

Multi-homogeneous polynomial systems arise in many applications. In [10] we provide bit complexity estimates for solving them 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 assumptions essentially imply that the Jacobian matrix of the system under study has maximal rank at the solution set and that this solution set if finite. 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.

6.2. Solving Systems over the Reals and Applications

6.2.1. Univariate real root isolation in an extension field and applications

In [11] we present algorithmic, complexity and implementation results for the problem of isolating the real roots of a univariate polynomial in $B_\alpha \in L[y]$, where $L = \mathbb{Q}(\alpha)$ is a simple algebraic extension of the rational numbers. We revisit two approaches for the problem. In the first approach, using resultant computations, we perform a reduction to a polynomial with integer coefficients and we deduce a bound of $\widetilde{\mathcal{O}}_B(N^8)$ for isolating the real roots of B_α, where N is an upper bound on all the quantities (degree and bitsize) of the input polynomials. The bound becomes $\widetilde{\mathcal{O}}_B(N^7)$ if we use Pan's algorithm for isolating the real roots. In the second approach we isolate the real roots working directly on the polynomial of the input. We compute improved separation bounds for the roots and we prove that they are optimal, under mild assumptions. For isolating the real roots we consider a modified Sturm algorithm, and a modified version of `descartes`' algorithm. For the former we prove a Boolean complexity bound of $\widetilde{\mathcal{O}}_B(N^{12})$ and for the latter a bound of $\widetilde{\mathcal{O}}_B(N^5)$. We present aggregate separation bounds and complexity results for isolating the real roots of all polynomials B_{α_k}, when

α_k runs over all the real conjugates of α. We show that we can isolate the real roots of all polynomials in $\widetilde{\mathcal{O}}_B(N^5)$. Finally, we implemented the algorithms in C as part of the core library of MATHEMATICA and we illustrate their efficiency over various data sets.

6.2.2. On the Maximal Number of Real Embeddings of Spatial Minimally Rigid Graphs

The number of embeddings of minimally rigid graphs in $\mathbb{R}^D$ is (by definition) finite, modulo rigid transformations, for every generic choice of edge lengths. Even though various approaches have been proposed to compute it, the gap between upper and lower bounds is still enormous. Specific values and its asymptotic behavior are major and fascinating open problems in rigidity theory. Our work in [13] considers the maximal number of real embeddings of minimally rigid graphs in $\mathbb{R}^3$. We modify a commonly used parametric semi-algebraic formulation that exploits the Cayley-Menger determinant to minimize the *a priori* number of complex embeddings, where the parameters correspond to edge lengths. To cope with the huge dimension of the parameter space and find specializations of the parameters that maximize the number of real embeddings, we introduce a method based on coupler curves that makes the sampling feasible for spatial minimally rigid graphs. Our methodology results in the first full classification of the number of real embeddings of graphs with 7 vertices in $\mathbb{R}^3$, which was the smallest open case. Building on this and certain 8-vertex graphs, we improve the previously known general lower bound on the maximum number of real embeddings in $\mathbb{R}^3$.

6.2.3. Lower bounds on the number of realizations of rigid graphs

Computing the number of realizations of a minimally rigid graph is a notoriously difficult problem. Towards this goal, for graphs that are minimally rigid in the plane, we take advantage of a recently published algorithm, which is the fastest available method, although its complexity is still exponential. Combining computational results with the theory of constructing new rigid graphs by gluing, in [4] we give a new lower bound on the maximal possible number of (complex) realizations for graphs with a given number of vertices. We extend these ideas to rigid graphs in three dimensions and we derive similar lower bounds, by exploiting data from extensive Gröbner basis computations.

6.2.4. The Complexity of Subdivision for Diameter-Distance Tests

In [1] we present a general framework for analyzing the complexity of subdivision-based algorithms whose tests are based on the sizes of regions and their distance to certain sets (often varieties) intrinsic to the problem under study. We call such tests diameter-distance tests. We illustrate that diameter-distance tests are common in the literature by proving that many interval arithmetic-based tests are, in fact, diameter-distance tests. For this class of algorithms, we provide both non-adaptive bounds for the complexity, based on separation bounds, as well as adaptive bounds, by applying the framework of continuous amortization. Using this structure, we provide the first complexity analysis for the algorithm by Plantinga and Vegeter for approximating real implicit curves and surfaces. We present both adaptive and non-adaptive a priori worst-case bounds on the complexity of this algorithm both in terms of the number of subregions constructed and in terms of the bit complexity for the construction. Finally, we construct families of hypersurfaces to prove that our bounds are tight.

6.2.5. Real root finding for equivariant semi-algebraic systems

Let R be a real closed field. In [19] we consider basic semi-algebraic sets defined by n-variate equations/inequalities of s symmetric polynomials and an equivariant family of polynomials, all of them of degree bounded by $2d < n$. Such a semi-algebraic set is invariant by the action of the symmetric group. We show that such a set is either empty or it contains a point with at most $2d-1$ distinct coordinates. Combining this geometric result with efficient algorithms for real root finding (based on the critical point method), one can decide the emptiness of basic semi-algebraic sets defined by s polynomials of degree d in time $(sn)^{O(d)}$. This improves the state-of-the-art which is exponential in n. When the variables $x_1, ..., x_n$ are quantified and the coefficients of the input system depend on parameters $y_1, ..., y_t$, one also demonstrates that the corresponding one-block quantifier elimination problem can be solved in time $(sn)^{O(dt)}$.

6.2.6. Exact algorithms for semidefinite programs with degenerate feasible set

Let $A_0, ..., A_nn$ be $m \times m$ symmetric matrices with entries in Q, and let $A(x)$ be the linear pencil $A_0 + x_1 A_1 + \cdots + x_n A_n$, where $x = (x1, ..., xn)$ are unknowns. The linear matrix inequality (LMI) $A(x) \succeq 0$ defines the subset of R^n, called spectrahedron, containing all points x such that $A(x)$ has nonnegative eigenvalues. The minimization of linear functions over spectrahedra is called semidefinite programming (SDP). Such problems appear frequently in control theory and real algebra, especially in the context of nonnegativity certificates for multivariate polynomials based on sums of squares. Numerical software for solving SDP are mostly based on the interior point method, assuming some non-degeneracy properties such as the existence of interior points in the admissible set. In [21], we design an exact algorithm based on symbolic homotopy for solving semidefinite programs without assumptions on the feasible set, and we analyze its complexity. Because of the exactness of the output, it cannot compete with numerical routines in practice but we prove that solving such problems can be done in polynomial time if either n or m is fixed.

6.2.7. A lower bound on the positive semidefinite rank of convex bodies

The positive semidefinite rank of a convex body C is the size of its smallest positive semidef-inite formulation. In [3] 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.

6.2.8. On the complexity of computing real radicals of polynomial systems

Let $f = (f_1, ..., f_s)$ be a sequence of polynomials in $Q[X_1, ..., X_n]$ of maximal degree D and $V \subset C^n$ be the algebraic set defined by f and r be its dimension. The real radical $\sqrt[re]{\langle f \rangle}$ associated to f is the largest ideal which defines the real trace of V. In [20] when V is smooth, we show that $\sqrt[re]{\langle f \rangle}$ has a finite set of generators with degrees bounded by V. Moreover, we present a probabilistic algorithm of complexity $(snDn)^{O(1)}$ to compute the minimal primes of $\sqrt[re]{\langle f \rangle}$. When V is not smooth, we give a probabilistic algorithm of complexity $s^{O(1)}(nD)^{O(nr2^r)}$ to compute rational parametrizations for all irreducible components of the real algebraic set $V \cap R^n$. Experiments are given to show the efficiency of our approaches.

6.2.9. Algorithms for Weighted Sums of Squares Decomposition of Non-negative Univariate Polynomials

It is well-known that every non-negative univariate real polynomial can be written as the sum of two polynomial squares with real coefficients. When one allows a weighted sum of finitely many squares instead of a sum of two squares, then one can choose all coefficients in the representation to lie in the field generated by the coefficients of the polynomial. In particular, this allows an effective treatment of polynomials with rational coefficients. In [9], we describe, analyze and compare both from the theoretical and practical points of view, two algorithms computing such a weighted sums of squares decomposition for univariate polynomials with rational coefficients. The first algorithm, due to the third author relies on real root isolation, quadratic approximations of positive polynomials and square-free decomposition but its complexity was not analyzed. We provide bit complexity estimates, both on the runtime and the output size of this algorithm. They are exponential in the degree of the input univariate polynomial and linear in the maximum bitsize of its complexity. This analysis is obtained using quantifier elimination and root isolation bounds. The second algorithm, due to Chevillard, Harrison, Joldes and Lauter, relies on complex root isolation and square-free decomposition and has been introduced for certifying positiveness of poly-nomials in the context of computer arithmetics. Again, its complexity was not analyzed. We provide bit complexity estimates, both on the runtime and the output size of this algorithm, which are polynomial in the degree of the input polynomial and linear in the maximum bitsize of its complexity. This analysis is obtained using Vieta's formula and root isolation bounds. Finally, we report on our implementations of both algorithms and compare them in practice on several application benchmarks. While the second algorithm is, as expected from the complexity result, more efficient on most of examples, we exhibit families of non-negative polynomials for which the first algorithm is better.

6.2.10. On Exact Polya and Putinar's Representations

We consider the problem of finding exact sums of squares (SOS) decompositions for certain classes of non-negative multivariate polynomials, relying on semidefinite programming (SDP) solvers. In [18] we start by providing a hybrid numeric-symbolic algorithm computing exact rational SOS decompositions for polynomials lying in the interior of the SOS cone. It computes an approximate SOS decomposition for a perturbation of the input polynomial with an arbitrary-precision SDP solver. An exact SOS decomposition is obtained thanks to the perturbation terms. We prove that bit complexity estimates on output size and runtime are both polynomial in the degree of the input polynomial and simply exponential in the number of variables. Next, we apply this algorithm to compute exact Polya and Putinar's representations respectively for positive definite forms and positive polynomials over basic compact semi-algebraic sets. We also compare the implementation of our algorithms with existing methods in computer algebra including cylindrical algebraic decomposition and critical point method.

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

6.3.1. Linear Repairing Codes and Side-Channel Attacks

To strengthen the resistance of countermeasures based on secret sharing, several works have suggested to use the scheme introduced by Shamir in 1978, which proposes to use the evaluation of a random d-degree polynomial into $nd+1$ public points to share the sensitive data. Applying the same principles used against the classical Boolean sharing, all these works have assumed that the most efficient attack strategy was to exploit the minimum number of shares required to rebuild the sensitive value; which is $d+1$ if the reconstruction is made with Lagrange's interpolation. In [2], we highlight first an important difference between Boolean and Shamir's sharings which implies that, for some signal-to-noise ratio, it is more advantageous for the adversary to observe strictly more than d + 1 shares. We argue that this difference is related to the existence of so-called exact linear repairing codes, which themselves come with reconstruction formulae that need (much) less information (counted in bits) than Lagrange's interpolation. In particular, this result implies that, contrary to what was believed, the choice of the public points in Shamir's sharing has an impact on the countermeasure strength. As another contribution, we exhibit a positive impact of the existence of linear exact repairing schemes; we indeed propose to use them to improve the state-of-the-art multiplication algorithms dedicated to Shamir's sharing. We argue that the improvement can be effective when the multiplication operation in the base field is at least two times smaller than in its sub-fields.

6.3.2. On the Use of Independent Component Analysis to Denoise Side-Channel Measurements

Independent Component Analysis (ICA) is a powerful technique for blind source separation. It has been successfully applied to signal processing problems, such as feature extraction and noise reduction , in many different areas including medical signal processing and telecommunication. In [17], we propose a framework to apply ICA to denoise side-channel measurements and hence to reduce the complexity of key recovery attacks. Based on several case studies, we afterwards demonstrate the overwhelming advantages of ICA with respect to the commonly used preprocessing techniques such as the singular spectrum analysis. Mainly, we target a software masked implementation of an AES and a hardware unprotected one. Our results show a significant Signal-to-Noise Ratio (SNR) gain which translates into a gain in the number of traces needed for a successful side-channel attack. This states the ICA as an important new tool for the security assessment of cryptographic implementations.

SECRET Project-Team

7. New Results

7.1. Symmetric cryptology

Participants: Xavier Bonnetain, Christina Boura, Anne Canteaut, Pascale Charpin, Daniel Coggia, Sébastien Duval, Gaëtan Leurent, María Naya Plasencia, Léo Perrin, Yann Rotella, André Schrottenloher, Ferdinand Sibleyras.

7.1.1. Block ciphers

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

- Nonlinear approximations of block ciphers: A. Canteaut, together with C. Beierle and G. Leander have exhibited the relationship between nonlinear invariants for block ciphers and nonlinear approximations. They have shown that, in some cases, the linear hull effect may be formalized in terms of nonlinear invariants. They have also introduced a new framework to study the probability of nonlinear approximations over iterated block ciphers [13], [26]
- Impossible differential cryptanalysis: C. Boura, V. Lallemand and M. Naya-Plasencia have introduced new techniques and complexity analyses for impossible differential cryptanalysis. They also showed that the technique of multiple differentials can be applied to impossible differential attacks [16]
- Construction of lightweight MDS matrices: S. Duval and G. Leurent have exhibited MDS matrices with the lowest known implementation cost. They have been constructed by a search through a space of circuits yielding MDS matrices [20], [11]

7.1.2. Stream ciphers

Stream ciphers provide an alternative to block-cipher-based encryption schemes. They are especially well-suited in applications which require either extremely fast encryption or a very low-cost hardware implementation.
Recent results:

- 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].
- Cryptanalysis of Goldreich pseudo-random generator: Goldreich's PRG is a theoretical construction which expands a short random string into a long pseudo-random string by applying a simple d-ary predicate to public random sized subsets of the bits of the seed. While the security of Goldreich's PRG has been thoroughly investigated, with a variety of results deriving provable security guarantees against classes of attacks in some parameter regimes and necessary criteria to be satisfied by the underlying predicate, little was known about its concrete security and efficiency. Motivated by the hope of getting practical instantiations of this construction, Y. Rotella and his co-authors initiated a study of the concrete security of Goldreich's PRG, and evaluated its resistance to cryptanalytic attacks. They developped a new guess-and-determine-style attack, and identified new criteria which captured the security guarantees [44].

7.1.3. Authenticated encryption

A limitation of all classical block ciphers is that they aim at protecting confidentiality only, while most applications need both encryption and authentication. These two functionalities are provided by using a block cipher like the AES together with an appropriate mode of operation. However, it appears that the most widely-used mode of operation for authenticated encryption, AES-GCM, is not very efficient for high-speed networks. Also, the security of the GCM mode completely collapses when an IV is reused. These severe drawbacks have then motivated an international competition named CAESAR, partly supported by the NIST, which has been launched in order to define some new authenticated encryption schemes [0]. The project-team is involved in a national cryptanalytic effort in this area led by the BRUTUS project funded by the ANR. In this context, the members of the project-team have obtained some cryptanalytic results on several candidates to the CAESAR competition.

Recent results:

- State-recovery attack on a simplified version of Ketje Jr. [21], [34]
- Cryptanalysis of Morus, one of the finalists of the CAESAR competition [42]

7.1.4. 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:

- 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 [15], [25]. 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.
- Boomerang Uniformity of Sboxes: The boomerang attack is a cryptanalysis technique against block ciphers which combines two differentials for the upper part and the lower part of the cipher. The Boomerang Connectivity Table (BCT) is a tool introduced by Cid *et al.* at Eurocrypt 2018 for analysing the dependency between these two differentials. C. Boura and A. Canteaut [14] have provided an in-depth analysis of BCT, by studying more closely differentially 4-uniform Sboxes. They have completely characterized the BCT of all differentially 4-uniform permutations of 4 bits and then study these objects for some cryptographically relevant families of Sboxes, as the inverse function and quadratic permutations. These two families are the first examples of differentially 4-uniform Sboxes optimal against boomerang attacks for an even number of variables, answering an open question raised by Cid *et al.*.
- CCZ equivalence of Sboxes: A. Canteaut and L. Perrin have characterized CCZ-equivalence as a property of the zeroes in the Walsh spectrum of an Sbox (or equivalently in their DDT). They used this framework to show how to efficiently upper bound the number of distinct EA-equivalence classes in a given CCZ-equivalence class. More importantly, they proved that CCZ-equivalence can be reduced to the association of EA-equivalence and an operation called twisting. They then revisited several results from the literature on CCZ-equivalence and showed how they can be interpreted in light of this new framework [18], [29]

[0] http://competitions.cr.yp.to/caesar.html

- Links between linear and differential properties of Sboxes: P. Charpin together with J. Peng has established new links between the differential uniformity and the nonlinearity of some Sboxes in the case of two-valued functions and quadratic functions. More precisely, they have exhibited a lower bound on the nonlinearity of monomial permutations depending on their differential uniformity, as well as an upper bound in the case of differentially two-valued functions [19], [55]
- Construction of building-blocks with resistance against fault-attacks at a low implementation overhead [50].

7.1.5. 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 provable security, and we know that their use is secure as long as the underlying primitive is secure, and we respect some limits on the amount of data processed. The analysis of generic attack helps us understand what happens when the hypotheses of the security proofs do not hold, or the corresponding limits are not respected. Comparing proofs and attacks also shows gaps where our analysis is incomplete, and when improved proof or attacks are required.
Recent results:

- Use of block ciphers operating on small blocks with the CTR mode [53]: the security proof of the CTR mode 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 less problematic than on CBC. However, G. Leurent and F. Sibleyras have exhibited concrete attacks against the CTR mode when processing close to $2^{n/2}$ blocks of data, and have shown that an attacker can actually extract as much information as in the case of CBC encryption.
- Generic attacks against some MAC constructions based on block ciphers [52]: G. Leurent and F. Sibleyras, together with M. Nandi, have studied the security of several recent MAC constructions with provable security beyond the birthday bound, namely `SUM-ECBC`, `PMAC+`, `3kf9`, `GCM-SIV2`, and some variants. They described a new cryptanalysis technique for double-block MACs and they showed how to build a forgery attack with query complexity $\mathcal{O}(2^{3n/4})$, proving that these schemes do not reach full security in the information-theoretic model. Surprisingly, their attack on `LightMAC+` invalidates a recent security proof by Naito. Moreover, they gave the first attack against `SUM-ECBC` and `GCM-SIV2`, with complexity below 2^n.

7.2. Code-based cryptography

Participants: Rodolfo Canto Torres, Thomas Debris, Matthieu Lequesne, Nicolas Sendrier, Jean-Pierre Tillich, Valentin Vasseur.

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.

Our recent work on code-based cryptography has to be seen in the context of the recently launched NIST competition whose purpose is to standardize quantum-safe public-key primitives. This call concerns all three major cryptographic primitives, namely public-key cryptosytems, key-exchange protocols and digital signature schemes. The most promising techniques today for addressing this issue are code-based cryptography, lattice-based cryptography, mutivariate cryptography, and hash-based cryptography.

Our contributions in this area are two-fold and consist in:

- designing and analysis new code-based solutions;
- cryptanalyzing code-based schemes, especially candidates to the NIST competition.

7.2.1. Design of new code-based solutions

The members of the project-team have submitted several candidates to the NIST competition, including a key-exchange protocol based on quasi-cyclic MDPC codes [41]. Their recent work on MDPC codes is important in this context in order to carefully analyze the properties of this candidate.

Recent results:

- Thwarting the GJS attack: the decryption algorithm of the QC-MDPC cryptosystem is based on an iterative bit-flipping algorithm, which fails with a small probability. These failures have been exploited in 2016 by Guo, Johansson and Stankovski to perform a key-recovery attack. JP Tillich recently analyzed how this attack can be avoided by increasing the key size of the scheme. Most notably, he proved that, under a very reasonable assumption, the error probability after decoding decays almost exponentially with the code-length with just two iterations of bit-flipping. With an additional assumption, it even decays exponentially with an unbounded number of iterations, implying that in this case the increase of the key size equired for resisting to the GJS attack is only moderate [54].
- Design of a new KEM with IND-CCA2 security in a model considering decoding failures [46]: M. Lequesne, N. Sendrier and their co-authors explored the underlying causes of the GJS attack, how it can be improved and how it can be mitigated. They derived a new timing attack performing well even in cases which were more challenging to the GJS attack. They also showed how to construct a new KEM, called ParQ that can reduce the decryption failure rate to a level negligible in the security parameter. They formally proved the IND-CCA2 security of ParQ, in a model that considers decoding failures.
- Design of a new code-based signature scheme [81]: T. Debris, N. Sendrier and JP Tillich recently proposed a "hash-and-sign" code-based signature scheme called `Wave`, which uses a family of ternary generalized (U, U + V) codes. `Wave` achieves existential unforgeability under adaptive-chosen-message attacks in the random oracle model with a tight reduction to two assumptions from coding theory: one is a distinguishing problem that is related to the trapdoor inserted in the scheme, the other one is a multiple-target version of syndrome decoding. This scheme enjoys efficient signature and verification algorithms. For 128-bit security, signature are 8000-bit long and the public-key size is slightly smaller than one megabyte.

7.2.2. Cryptanalysis of code-based schemes

Recent results:

- Cryptanalysis of two public-key cryptosystems based on the rank syndrome decoding problem [41]: JP Tillich and his co-authors proposed an attack on the Rank Syndrome Decoding problem which improves the previously best known algorithm for solving this problem. This attack breaks for some parameters some recently proposed cryptosystems based on LRPC codes or Gabidulin codes, including Loidreau's cryptosystem and the LRPC cryptosystem.
- Cryptanalysis of the NIST submission `RankSign` and of a recently proposed IBE scheme: T. Debris and JP Tillich have presented an algebraic attack against `RankSign` that exploits the fact that the augmented LRPC codes used in this scheme have codewords with a very low weight. This attack shows that all the parameters proposed for this candidate can be broken. They also proved that, for the IBE scheme based on `RankSign`, the problem is deeper than finding a new signature in rank-based cryptography, since they found an attack on the generic problem upon which the security reduction relies [45].

- Cryptanalysis of the EDON-K key encapsulation mechanism submitted to the NIST competition: EDON-K is a candidate to the NIST competition which is inspired by the McEliece scheme but uses another family of codes defined over $\mathbb{F}_{2^{128}}$ instead of $\mathbb{F}_2$ and is not based on the Hamming metric. M. Lequesne and JP Tillich presented an attack making the scheme insecure for the intended use. Indeed, recovering the error in the McEliece scheme corresponding to EDON-K can be viewed as a decoding problem for the rank-metric with a super-code of an LRPC code of very small rank A suitable parity-check matrix for this super-code can then be easily derived from the public key and used to recover the error [51].
- Attack against RLCE [80]: M. Lequesne and JP Tillich, together with A. Couvreur, recently presented a key-recovery attack against the Random Linear Code Encryption (RLCE) scheme recently submitted by Y. Wang to the NIST competition. This attack recovers the secret-key for all the short key-parameters proposed by the author. It uses a polynomial-time algorithm based on a square code distinguisher.

7.3. Quantum Information

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

Two PhD students within the project-team work on this topic. First, Antoine Grospellier, co-advised by A. Leverrier and O. Fawzi (Ens Lyon), studies efficient decoding algorithms for quantum LDPC codes. Beyond their intrinsic interest for channel-coding problems, such algorithms would be particularly relevant in the context of quantum fault-tolerance, since they would allow to considerably reduce the required overhead to obtain fault-tolerance in quantum computation. Vivien Londe is co-advised by A. Leverrier and G. Zémor (IMB) and his thesis is devoted to the design of better quantum LDPC codes: the main idea is to generalize the celebrated toric code of Kitaev by considering cellulations of manifolds in higher dimensions. A recent surprising result was that this approach leads to a much better behaviour than naively expected and a major challenge is to explore the mathematics behind this phenomenon in order to find even better constructions, or to uncover potential obstructions.

Recent results:

- Decoding algorithm for quantum expander codes [48], [47], [56] In this work, A. Grospellier, A. Leverrier and O. Fawzi analyze an efficient decoding algorithm for quantum expander codes and prove that it can correct a linear number of random errors with a negligible failure probability. 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.

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. The interest of continuous variables for quantum cryptography was recently recognized by being awarded a 10 M€ funding from the Quantum Flagship and SECRET will contribute to this project by studying the security of new key distribution protocols [88].

Recent results:

- Security proof for two-way continuous-variable quantum key distribution [22]: while many quantum key distribution protocols are one-way in the sense that quantum information is sent from one party to the other, it can be beneficial in terms of performance to consider two-way protocols where the quantum states perform a round-trip between the two parties. In this paper (to appear in *Physical Review A*), we show how to exploit the symmetries of the protocols in phase-space to establish their security against the most general attacks allowed by quantum theory.
- Investigating the optimality of ancilla-assisted linear optical Bell measurements [24]: Due to its experimental and theoretical simplicity, linear quantum optics has proved to be a promising route for the early implementation of important quantum communication protocols. A. Olivo and F. Grosshans study the efficiency of non ambiguous Bell measurements in this model and show both theoretical and numerical bounds depending on the number of ancilla qubits. This is important for understanding what resources are needed for building quantum repeaters, the last missing building block for secure long distance quantum key distribution networks.

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 commitment and zero-knowledge proofs [30]: A. Chailloux and A. Leverrier constructed 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. R. Bricout and A. Chailloux also studied relativistic multi-round bit commitment schemes. They showed optimal classical cheating strategies for the canonical F_Q commitment scheme.

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.

Recent results:

- Hidden-shift quantum cryptanalysis [43]: X. Bonnetain and M. Naya-Plasencia have obtained new results that consider the tweak proposed at Eurocrypt 2017 of using modular additions to counter Simon's attacks. They have developed new algorithms that improve and generalize Kuperberg's algorithm for the hidden shift problem. Thanks to their improved algorithm, they have been able to build a quantum attack in the superposition model on Poly1305, proposed at FSE 2005, largely used and claimed to be quantumly secure. They also analyzed the security of some classical symmetric constructions with concrete parameters, to evaluate the impact and practicality of the proposed tweak, concluding that it does not seem to be efficient
- Quantum algorithm for the k-XOR problem [49]: The k-XOR (or generalized birthday) problem aims at finding k elements of n-bits, drawn at random, such that the XOR of all of them is 0. The algorithms proposed by Wagner more than 15 years ago remain the best known classical algorithms for solving it, when disregarding logarithmic factors. M. Naya-Plasencia and A. Schrottenloher, together with L. Grassi, studied this problem in the quantum setting and provided algorithms with the best known quantum time-complexities. In particular, they were able to considerably improve the 3-XOR algorithm.
- Quantum cryptanalysis of CSIDH and Ordinary Isogeny-based Schemes [68]: CSIDH is a recent proposal by Castryck et al. for post-quantum non-interactive key-exchange. It is similar in design to a scheme by Couveignes, Rostovtsev and Stolbunov, but it replaces ordinary elliptic curves by supersingular elliptic curves. Although CSIDH uses supersingular curves, it can attacked by a quantum subexponential hidden shift algorithm due to Childs et al. While the designers of CSIDH claimed that the parameters they suggested ensures security against this algorithm, X. Bonnetain and A. Schrottenloher showed that these security parameters were too optimistic: they improved the hidden shift algorithm and gave a precise complexity analysis in this context, which greatly reduced the complexity. For example, they showed that only 2^{35} quantum equivalents of a key-exchange are sufficient to break the 128-bit classical, 64-bit quantum security parameters proposed, instead of 2^{62}. They also extended their analysis to ordinary isogeny computations, and showed that an instance proposed by De Feo, Kieffer and Smith and expected to offer 56 bits of quantum security can be broken in 2^{38} quantum evaluations of a key exchange.

SPECFUN Project-Team

7. New Results

7.1. Computing solutions of linear Mahler equations

Mahler equations relate evaluations of the same function f at iterated bth powers of the variable. They arise in particular in the study of automatic sequences and in the complexity analysis of divide-and-conquer algorithms. Recently, the problem of solving Mahler equations in closed form has occurred in connection with number-theoretic questions. A difficulty in the manipulation of Mahler equations is the exponential blow-up of degrees when applying a Mahler operator to a polynomial. In [3], Frédéric Chyzak and Philippe Dumas, together with Thomas Dreyfus (IRMA, Université de Strasbourg) and Marc Mezzarobba (external collaborator from Sorbonne Université), have presented algorithms for solving linear Mahler equations for series, polynomials, and rational functions, and have obtained polynomial-time complexity under a mild assumption. The article was formally accepted and published this year.

7.2. Becker's conjecture on Mahler functions

In 1994, Becker conjectured that if $F(z)$ is a k-regular power series, then there exists a k-regular rational function $R(z)$ such that $F(z)/R(z)$ satisfies a Mahler-type functional equation with polynomial coefficients where the initial coefficient satisfies $a_0(z) = 1$. In [1], Frédéric Chyzak and Philippe Dumas, together with Jason P. Bell (University of Waterloo, Canada) and Michael Coons (University of Newcastle, Australia) have proved Becker's conjecture in the best-possible form: they have shown that the rational function $R(z)$ can be taken to be a polynomial $z^{\gamma}Q(z)$ for some explicit non-negative integer γ and such that $1/Q(z)$ is k-regular. The article was formally accepted this year.

7.3. Generalized Hermite reduction, creative telescoping and definite integration of D-finite functions

Hermite reduction is a classical algorithmic tool in symbolic integration. It is used to decompose a given rational function as a sum of a function with simple poles and the derivative of another rational function. Alin Bostan, Frédéric Chyzak, and Pierre Lairez, together with Bruno Salvy (project-team AriC) have extended Hermite reduction to arbitrary linear differential operators instead of the pure derivative. They have also developped efficient algorithms for this reduction, and then applied the generalized Hermite reduction to the computation of linear operators satisfied by single definite integrals of D-finite functions of several continuous or discrete parameters. The resulting algorithm is a generalization of reduction-based methods for creative telescoping. Their article [6] was published at the ISSAC conference.

7.4. Bijections between Łukasiewicz walks and generalized tandem walks

In [9], Frédéric Chyzak, together with Karen Yeats (University of Waterloo, Canada), have studied the enumeration by length of several walk models on the square lattice. They have obtained bijections between walks in the upper half-plane returning to the x-axis and walks in the quarter plane. An ongoing work by Bostan, Chyzak, and Mahboubi has given a bijection for models using small north, west, and south-east steps. The work in [9] has adapted and generalized it to a bijection between half-plane walks using those three steps in two colours and a quarter-plane model over the symmetrized step set consisting of north, north-west, west, south, south-east, and east. They have then generalized their bijections to certain models with large steps: for given $p \geq 1$, a bijection has been given between the half-plane and quarter-plane models obtained by keeping the small south-east step and replacing the two steps north and west of length 1 by the $p + 1$ steps of length p in directions between north and west. An article was submitted this year.

7.5. Putting Fürer's algorithm into practice with the BPAS library

Fast algorithms for integer and polynomial multiplication play an important role in scientific computing as well as in other disciplines. In 1971, Schönhage and Strassen designed an algorithm that improved the multiplication time for two integers of at most n bits to $O(\log n \log \log n)$. Martin Fürer presented a new algorithm that runs in $O(n \log n \cdot 2^{O(\log^* n)})$, where $\log^* n$ is the iterated logarithm of n. In a submitted article, Svyatoslav Covanov, together with Davood Mohajerani, Marc Moreno Maza and Lin-Xiao Wang, have explained how one can put Fürer's ideas into practice for multiplying polynomials over a prime field $\mathbb{Z}/p\mathbb{Z}$, for which p is a Generalized Fermat prime of the form $p = r^k + 1$ where k is a power of 2 and r is of machine word size. When k is at least 8, they have shown that multiplication inside such a prime field can be efficiently implemented via Fast Fourier Transform (FFT). Taking advantage of Cooley-Tukey tensor formula and the fact that r is a $2k$-th primitive root of unity in $\mathbb{Z}/p\mathbb{Z}$, they have obtained an efficient implementation of FFT over $\mathbb{Z}/p\mathbb{Z}$. This implementation outperforms comparable implementations either using other encodings of $\mathbb{Z}/p\mathbb{Z}$ or other ways to perform multiplication in $\mathbb{Z}/p\mathbb{Z}$.

7.6. Fast coefficient computation for algebraic power series in positive characteristic

In [5], Alin Bostan and Philippe Dumas, together with Xavier Caruso (CNRS, Rennes) and Gilles Christol (IMJ, Paris) have studied the algorithmic question of coefficient computation of algebraic power series in positive characteristic. They revisited Christol's theorem on algebraic power series in positive characteristic and proposed another proof for it. Their new proof combines several ingredients and advantages of existing proofs, which make it very well-suited for algorithmic purposes. The construction used in the new proof was then applied to the design of a new efficient algorithm for computing the Nth coefficient of a given algebraic power series over a perfect field of characteristic p. This algorithm has several nice features: it is more general, more natural and more efficient than previous algorithms. Not only the arithmetic complexity of the new algorithm is linear in $\log N$ and quasi-linear in p, but its dependency with respect to the degree of the input is much smaller than in the previously best algorithm. Moreover, when the ground field is finite, the new approach yields an even faster algorithm, whose bit complexity is linear in $\log N$ and quasi-linear in $\sqrt{p}$.

7.7. Counting walks with large steps in an orthant

In the past fifteen years, the enumeration of lattice walks with steps taken in a prescribed set and confined to a given cone, especially the first quadrant of the plane, has been intensely studied. As a result, the generating functions of quadrant walks are now well-understood, provided the allowed steps are *small*. In particular, having small steps is crucial for the definition of a certain group of bi-rational transformations of the plane. It has been proved that this group is finite if and only if the corresponding generating function is D-finite. This group is also the key to the uniform solution of 19 of the 23 small step models possessing a finite group. In contrast, almost nothing was known for walks with arbitrary steps. In [7], Alin Bostan together with Mireille Bousquet-Mélou (CNRS, Bordeaux) and Stephen Melczer (U. Pennsylvania, Philadelphia, USA), extended the definition of the group, or rather of the associated orbit, to this general case, and generalized the above uniform solution of small step models. When this approach works, it invariably yields a D-finite generating function. They applied it to many quadrant problems, including some infinite families. After developing the general theory, the authors of [7] considered the 13 110 two-dimensional models with steps in $\{-2, -1, 0, 1\}^2$ having at least one -2 coordinate. They proved that only 240 of them have a finite orbit, and solve 231 of them with our method. The 9 remaining models are the counterparts of the 4 models of the small step case that resist the uniform solution method (and which are known to have an algebraic generating function). They conjecture D-finiteness for their generating functions (but only two of them are likely to be algebraic!), and proved non-D-finiteness for the 12 870 models with an infinite orbit, except for 16 of them.

7.8. Subresultants of $(x - \alpha)^m$ and $(x - \beta)^n$, Jacobi polynomials and complexity

A previous article in 2017 described explicit expressions for the coefficients of the order-d polynomial subresultant of $(x-\alpha)^m$ and $(x-\beta)^n$ with respect to Bernstein's set of polynomials $\{(x-\alpha)^j(x-\beta)^{d-j},\, 0 \le j \le d\}$, for $0 \le d < \min\{m, n\}$. In [8], Alin Bostan, together with T. Krick, M. Valdettaro (U. Buenos Aires, Argentina) and A. Szanto (U. North Carolina, Raleigh, USA) further developed the study of these structured polynomials and showed that the coefficients of the subresultants of $(x-\alpha)^m$ and $(x-\beta)^n$ with respect to the monomial basis can be computed in *linear* arithmetic complexity, which is faster than for arbitrary polynomials. The result is obtained as a consequence of the amazing though seemingly unnoticed fact that these subresultants are scalar multiples of Jacobi polynomials up to an affine change of variables.

7.9. A numerical transcendental method in algebraic geometry

In “A transcendental method in algebraic geometry”, Griffiths emphasized the role of certain multivariate integrals, known as *periods*, “to construct a continuous invariant of arbitrary smooth projective varieties”. Periods often determine the projective variety completely and therefore its algebraic invariants. Translating periods into discrete algebraic invariants is a difficult problem, exemplified by the long standing Hodge conjecture which describes how periods determine the algebraic cycles within a projective variety.

Recent progress in computer algebra makes it possible to compute periods with high precision and put transcendental methods into practice. In [10], Pierre Lairez and Emre Sertöz focus on algebraic surfaces and give a numerical method to compute Picard groups. As an application, they count smooth rational curves on quartic surfaces using the Picard group. It is the first time that this kind of computation is performed.

CAIRN Project-Team

7. New Results

7.1. Reconfigurable Architecture Design

7.1.1. Algorithmic Fault Tolerance for Timing Speculative Hardware

Participants: Thibaut Marty, Tomofumi Yuki, Steven Derrien.

Timing speculation, also known as overclocking, is a well known approach to increase the computational throughput of processors and hardware accelerators. When used aggressively, timing speculation can lead to incorrect/corrupted results. As reported in the literature, timing errors can cause large numerical errors in the computation, and such occasional large errors can have devastating effect on the final output. The frequency of such errors depends on a number of factors, including the intensity of overclocking, operating temperature, voltage drops, variability within and across boards, input data, and so on. This makes it extremely difficult to determine a "safe" overclocking speed analytically or empirically. Several circuit-level error mitigating techniques have been proposed, but they are difficult to implement in modern FPGAs, and often involve significant area overhead. Instead of resorting to circuit level technique, we propose to rely on light-weight algorithm-level error detections techniques. This allows us to augment accelerators with low overhead mechanism to protect against timing errors, enabling aggressive timing speculation. We have demonstrated in [36] the validity of our approach for convolutional neural networks, where we use overclocking for the convolution stages. Our prototype on ZC706 demonstrated 68-77% computational throughput with negligible (<1%) area overhead.

7.1.2. 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 with 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 [43], [27] show that, compared to a software implementation, our approach offers speed-up by 8× while consuming 18× less energy. We have also shown how our approach 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. Our first experimental results have shown that this approach leads to best-case performance and energy efficiency when compared against static VLIW configurations [42].

7.1.3. 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 of heterogeneous multicore architectures using the HAsim/LEAP infrastructure. Two aspects are currently under development. The first one concerns the deployment of simulator models on the hybrid Xeon CPU-Arria 10 FPGA Intel platforms. The second one concerns the definition of simulation models of hardware accelerators. The core processor brick is a RISCV core.

7.1.4. Dynamic Fault-Tolerant Scheduling onto 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. Temporal and/or spatial redundancy is currently used to tackle this issue. 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 several run-time algorithms based on the primary/backup approach which is commonly used for its minimal resources utilization and high reliability [31], [30]. The aim of our work was to reduce the complexity of the algorithm in order to target real-time embedded systems without sacrificing reliability. This work is done in collaboration with Oliver Sinnen, PARC Lab., the University of Auckland.

7.1.5. 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 [21] 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.

7.1.6. Energy Constrained and Real-Time Scheduling and Assignment on Multicores

Participants: Olivier Sentieys, Angeliki Kritikakou, Lei Mo.

Multicore architectures have been used to enhance computing capabilities, but the energy consumption is still an important concern. Embedded application domains usually require less accurate, but always in-time, results. Imprecise Computation (IC) can be used to divide a task into a mandatory subtask providing a baseline Quality-of-Service (QoS) and an optional subtask that further increases the baseline QoS. Combining dynamic voltage and frequency scaling, task allocation and task adjustment, we can maximize the system QoS under real-time and energy supply constraints. However, the nonlinear and combinatorial nature of this problem makes it difficult to solve. In [25], we formulate a Mixed-Integer Non-Linear Programming (MINLP) problem to concurrently carry out task-to-processor allocation, frequency-to-task assignment and optional task adjustment. We provide a Mixed-Integer Linear Programming (MILP) form of this formulation without performance degradation and we propose a novel decomposition algorithm to provide an optimal solution with reduced computation time compared to state-of-the-art optimal approaches (22.6% in average). We also propose a heuristic version that has negligible computation time. In [24], we focus on QoS maximizing for dependent IC-tasks under real-time and energy constraints. Compared with existing approaches, we consider the joint-design problem, where task-to-processor allocation, frequency-to-task assignment, task scheduling and task adjustment are optimized simultaneously. The joint-design problem is formulated as an NP-hard Mixed-Integer Non-Linear Programming and it is safely transformed to a Mixed-Integer Linear Programming (MILP) without performance degradation. Two methods (basic and accelerated version) are proposed to find the optimal solution to MILP problem. They are based on problem decomposition and provide a controllable way to trade-off the quality of the solution and the computational complexity. The optimality of the proposed methods is proved rigorously, and the experimental results show reduced computation time (23.7% in average) compared with existing optimal methods. Finally, in [50] we summarize the problem and the methods for imprecise computation task mapping on multicore Wireless Sensor Networks.

7.1.7. Real-Time Energy-Constrained Scheduling in Wireless Sensor and Actuator Networks

Participants: Angeliki Kritikakou, Lei Mo.

Wireless Sensor and Actuator Networks (WSANs) are emerging as a new generation of Wireless Sensor Networks (WSNs). Due to the coupling between the sensing areas of the sensors and the action areas of the actuators, the efficient coordination among the nodes is a great challenge. In our work in [23] we address the problem of distributed node coordination in WSANs aiming at meeting the user's requirements on the states of the Points of Interest (POIs) in a real-time and energy-efficient manner. The node coordination problem is formulated as a non-linear program. To solve it efficiently, the problem is divided into two correlated subproblems: the Sensor-Actuator (S-A) coordination and the Actuator-Actuator (A-A) coordination. In the S-A coordination, a distributed federated Kalman filter-based estimation approach is applied for the actuators to collaborate with their ambient sensors to estimate the states of the POIs. In the A-A coordination, a distributed Lagrange-based control method is designed for the actuators to optimally adjust their outputs, based on the estimated results from the S-A coordination. The convergence of the proposed method is proved rigorously. As the proposed node coordination scheme is distributed, we find the optimal solution while avoiding high computational complexity. The simulation results also show that the proposed distributed approach is an efficient and practically applicable method with reasonable complexity. In addition, the design of fast and effective coordination among sensors and actuators in Cyber-Physical Systems (CPS) is a fundamental, but challenging issue, especially when the system model is a priori unknown and multiple random events can simultaneously occur. In [37], we propose a novel collaborative state estimation and actuator scheduling algorithm with two phases. In the first phase, we propose a Gaussian Mixture Model (GMM)-based method using the random event physical field distribution to estimate the locations and the states of events. In the second phase, based on the number of identified events and the number of available actuators, we study two actuator scheduling scenarios and formulate them as Integer Linear Programming (ILP) problems with the objective to minimize the actuation delay. We validate and demonstrate the performance of the proposed scheme through both simulations and physical experiments for a home temperature control application.

7.1.8. 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. By including this parameter in the operating system, it is then possible to develop some 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 including P identical cores. We assumed that they can be reconfigured dynamically by authorizing the addition and/or removal of periodic tasks and each core schedules its local tasks by using the EDF algorithm [20]. This work is extended to address heterogeneous systems for which each task has different execution parameters. The objective of this extension work is to develop a new strategy for mapping N tasks to P heterogeneous cores of a given distributed system [32]. For these two architectures models, we formulated the problem as an Integer Linear Program (ILP) optimization problem. Assuming that the energy consumed by the communication is dependent on the distance between cores, we proposed a mapping strategy to minimize the total cost of communication between cores 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-core architecture case. Experimental results reveal the effectiveness of the proposed strategy by comparing the derived heuristics with the optimal ones, obtained by solving an ILP problem

7.1.9. Improving the Reliability of Wireless NoC

Participants: Olivier Sentieys, Joel Ortiz Sosa.

Wireless Network-on-Chip (WiNoC) is one of the most promising solutions to overcome multi-hop latency and high power consumption of modern many/multi core System-on- Chip (SoC). However, the design of efficient wireless links faces challenges to overcome multi-path propagation present in realistic WiNoC channels. In order to alleviate such channel effect, we propose a Time-Diversity Scheme (TDS) to enhance the reliability of on-chip wireless links using a semi-realistic channel model in [45]. First, we study the significant performance degradation of state-of-the-art wireless transceivers subject to different levels of multi-path propagation. Then we investigate the impact of using some channel correction techniques adopting standard performance metrics. Experimental results show that the proposed Time-Diversity Scheme significantly improves Bit Error Rate (BER) compared to other techniques. Moreover, our TDS allows for wireless communication links to be established in conditions where this would be impossible for standard transceiver architectures. Results on the proposed complete transceiver, designed using a 28-nm FDSOI technology, show a power consumption of 0.63mW at 1.0V and an area of 317 μm^2. Full channel correction is performed in one single clock cycle.

7.1.10. Optical Network-on-Chip (ONoC) for 3D Multiprocessor Architectures

Participants: Jiating Luo, Van Dung Pham, Cédric Killian, Daniel Chillet, Olivier Sentieys.

Photonics on silicon is now a technology that offers real opportunities in the context of multiprocessor interconnect. 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 latency. 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. We formulated the crosstalk noise and latency models and then proposed a Wavelength Allocation (WA) method in a ring-based WDM ONoC to reach performance and energy trade-offs based on application constraints. We show that for a 16-cluster ONoC architecture using 12 wavelengths, more than 10^5 allocation solutions exist and only 51 are on a Pareto front giving a tradeoff between latency and energy per bit derived from the BER. These optimized solutions reduce the execution time of the application by 37% and the energy from 7.6fJ/bit to 4.4fJ/bit. In

[22], we define high-level mechanisms which can handle wavelength allocation protocol of the communication medium for each data transfer between tasks. 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. We produce the communication configurations which are defined by the number of wavelengths for each communication, the level of quality for the communications, and the laser power levels. In [35], we define an Optical-Network-Interface (ONI) to connect a cluster of processors to the optical communication medium. This interface, constrained by the 10 Gb/s data-rate of the lasers, integrates Error Correcting Codes (ECC), laser drivers, and a communication manager. The ONI can select, at run-time, the communication mode to use depending on performance, latency or power constraints. The use of ECC is based on redundant bits which increases the transmission time, but saves power for a given Bit Error Rate (BER). Furthermore, the use of several wavelengths in parallel reduces latency and increases bandwidth, but also increases communication loss.

7.2. Compilation and Synthesis for Reconfigurable Platform

7.2.1. Compile Time Simplification of Sparse Matrix Code Dependences

Participant: Tomofumi Yuki.

Analyzing array-based computations to determine data dependences is useful for many applications including automatic parallelization, race detection, computation and communication overlap, verification, and shape analysis. For sparse matrix codes, array data dependence analysis is made more difficult by the use of index arrays that make it possible to store only the nonzero entries of the matrix (e.g., in $A[B[i]]$, B is an index array). Here, dependence analysis is often stymied by such indirect array accesses due to the values of the index array not being available at compile time. Consequently, many dependences cannot be proven unsatisfiable or determined until runtime. Nonetheless, index arrays in sparse matrix codes often have properties such as monotonicity of index array elements that can be exploited to reduce the amount of runtime analysis needed. In this work, we contribute a formulation of array data dependence analysis that includes encoding index array properties as universally quantified constraints. This makes it possible to leverage existing SMT solvers to determine whether such dependences are unsatisfiable and significantly reduces the number of dependences that require runtime analysis in a set of eight sparse matrix kernels. Another contribution is an algorithm for simplifying the remaining satisfiable data dependences by discovering equalities and/or subset relationships. These simplifications are essential to make a runtime-inspection-based approach feasible.

7.2.2. Automatic Parallelization Techniques for Time-Critical Systems

Participants: Steven Derrien, Mickael Dardaillon.

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 lend 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 [34]. Our current research direction aims [41] at studying the impact of compiler optimization on WCET estimates, and develop specific WCET aware compiler optimization flows. More specifically, we explore the use of iterative compilation (WCET-directed program optimization to explore the optimization space), with the objective to (i) allow flow facts to be automatically found and (ii) select optimizations that result in the lowest WCET estimates. We also explore to which extent code outlining helps, by allowing the selection of different optimization options for different code snippets of the application.

7.2.3. Design of High Throughput Mathematical Function Evaluators

Participant: Silviu Ioan Filip.

The evaluation of mathematical functions is a core component in many computing applications and has been a core topic in computer arithmetic since the inception of the field. In [28], we proposed an automatic method for the evaluation of functions via polynomial or rational approximations and its hardware implementation, on FPGAs. These approximations are evaluated using Ercegovac's iterative E-method adapted for FPGA implementation. The polynomial and rational function coefficients are optimized such that they satisfy the constraints of the E-method. It allows for an effective way to perform design space exploration when targeting high throughput.

7.2.4. Robust Tools for Computing Rational Chebyshev Approximations

Participant: Silviu Ioan Filip.

Rational functions are useful in a plethora of applications, including digital signal processing and model order reduction. They are nevertheless known to be much harder to work with in a numerical context than other, potentially less expressive families of approximating functions, like polynomials. In [19] we have proposed the use of a numerically robust way of representing rational functions, the barycentric form (*i.e.,* a ratio of partial fractions sharing the same poles). We use this form to develop scalable iterative algorithms for computing rational approximations to functions which minimize the uniform norm error. Our results are shown to significantly outperform previous state of the art approaches.

CAMUS Team

7. New Results

7.1. AutoParallel: A Python module for automatic parallelization and distributed execution of affine loop nests

Participant: Philippe Clauss.

The last improvements in programming languages, programming models, and frameworks have focused on abstracting the users from many programming issues. Among others, recent programming frameworks include simpler syntax, automatic memory management and garbage collection, which simplifies code re-usage through library packages, and easily configurable tools for deployment. For instance, Python has risen to the top of the list of the programming languages due to the simplicity of its syntax, while still achieving a good performance even being an interpreted language. Moreover, the community has helped to develop a large number of libraries and modules, tuning the most commonly used to obtain great performance.

However, there is still room for improvement when preventing users from dealing directly with distributed and parallel computing issues. This work proposes AutoParallel, a Python module to automatically find an appropriate task-based parallelization of affine loop nests to execute them in parallel in a distributed computing infrastructure. This parallelization can also include the building of data blocks to increase task granularity in order to achieve a good execution performance. Moreover, AutoParallel is based on sequential programming and only contains a small annotation in the form of a Python decorator so that anyone with little programming skills can scale up an application to hundreds of cores.

This work has been published in [18] and is the result of a collaboration between Philippe Clauss, Cristian Ramon-Cortes, PhD student, and Rosa M. Badia, his PhD advisor, both from the Barcelona Supercomputing Center, Spain.

7.2. Optimization of recursive functions by transformation into loops

Participants: Salwa Kobeissi, Philippe Clauss.

Recursion is a fundamental computing concept that offers the opportunity to elegantly solve various kinds of problems, particularly those whose solutions depend on solutions of smaller instances of their own. Nevertheless, today in imperative languages, recursive functions are still not considered sufficiently time-efficient in comparison with the alternative equivalent iterative code. Although many advanced and aggressive optimizers have been developed to enhance the performance of iterative control structures, there are still no such sophisticated and advanced techniques built for the sake of optimizing recursions.

We propose an approach that makes possible applying powerful optimizations on recursive functions through transforming them into loops. We are particularly interested in applying polyhedral optimization techniques which usually tackle affine loops. Therefore, the scope of our study is restricted to recursive functions whose control flow and memory accesses exhibit an affine behavior, which means that there exists a semantically equivalent affine loop nest, candidate for polyhedral optimizations. Accordingly, our approach is based on analyzing early executions of a recursive program using a Nested Loop Recognition algorithm, performing the convenient recursion-to-iteration transformation of the original program and, finally, applying further loop optimizations using the polyhedral compiler Polly. This approach brings recursion optimization techniques into a higher level in addition to widening the scope of the polyhedral model to include originally non-loop programs.

This work is the topic of Salwa Kobeissi's PhD. A first paper has been submitted to an international workshop.

7.3. Impact Study of Data Locality on Task-Based Applications Through the Heteroprio Scheduler

Participant: Bérenger Bramas.

Task-based parallelization is massively used in high-performance computing on heterogeneous hardware because it allows programmers to finely describe the intrinsic parallelism of the algorithms while ignoring the hardware details. However, this approach delegates the main decisions to the scheduler, making it a critical component responsible for the distribution of the tasks on the different types of processing unit. In a former work, Bérenger Bramas has proposed the Heteroprio scheduler, which has demonstrated to be extremely efficient in the computation of the fast multipole method or linear algebra factorizations/decompositions. However, the original version was not taking into account data locality leading to loss of execution efficiency from important data movements between the memory nodes.

The current work aimed at improving the Heteroprio scheduler by making it locality sensitive. The idea is to divide the task-lists to have as many lists as there are memory nodes. Then, the two main issues are to find where to store the new ready tasks and to decide how to iterate over all the task-lists. For the first problem, we have studied different locality scores to find the best memory node for each task, and we have demonstrated that taking into account the type of data access - read or write - allows for significant improvement. Concerning the iteration order, we have proposed to use a priority distance and a memory distance such that the tasks are stolen from memory nodes that are close but also that have opposite priorities.

All these ideas were implemented in a scheduler inside StarPU and have been validated on two applications: QrMumps from Alfredo Buttari (IRIT) and SpLDLT from Florent Lopez (Rutherford Appleton Laboratory, UK.). The performance study demonstrated the benefit of our approach with a significant improvement in terms of execution time and data movement. The executions were accelerated by 30% for QrMumps and 80% for SpLDLT. The results will now be written into a dedicated paper for publication.

7.4. Combining Locking and Data Management Interfaces

Participants: Jens Gustedt, Maxime Mogé, Mariem Saied, Daniel Salas.

7.4.1. *Ordered Read-Write Locks*

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 [2] 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 6.6 . In previous work it has demonstrated its efficiency for a large variety of platforms.

In the context of the thesis of Mariem Saied, a new domain specific language (DSL) has been completed that largely eases the implementation of applications with ORWL. In its first version it provides an interface for stencil codes. The approach allows to describe stencil codes quickly and efficiently, and leads to substantial speedups.

In the framework of the ASNAP project (see 9.1.2) we have used ordered read-write locks (ORWL) as a model to dynamically schedule a pipeline of parallel tasks that realize a parallel control flow of two nested loops; an outer *iteration* loop and an inner *data traversal* loop. Other than dataflow programming we emphasize on upholding the sequential modification order of each data object. As a consequence the visible side effects on any object can be guaranteed to be identical to a sequential execution. Thus the set of optimizations that are performed are compatible with C's abstract state machine and compilers could perform them, in principle, automatically and unobserved. See [19] for first results.

In the context of the Prim'Eau project (see 9.1.1) we use ORWL to integrate parallelism into an already existing `Fortran` application that computes floods in the region that is subject to the study. A first step of such a parallelization has been started by using ORWL on a process level. Our final goal will be to extend it to the thread level and to use the application structure for automatic placement on compute nodes.

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.

7.4.2. Low level locks

Our low level locks algorithm that is based on atomics and Linux' futexes [25] [26] has been integrated into the `musl` C library (see Section 6.7) and is thus deployed in several Linux distributions that use musl as their base.

7.5. 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. On recent multi-core hardware, performance of this code is often limited by memory bandwidth. We describe a multi-core PIC algorithm that achieves close-to-minimal number of memory transfers with the main memory, while at the same time exploiting SIMD instructions for numerical computations and exhibiting a high degree of OpenMP-level parallelism. Our algorithm keeps particles sorted by cell at every time step, and represents particles from the same cell using a linked list of fixed-capacity arrays, called chunks. Chunks support either sequential or atomic insertions, the latter being used to handle fast-moving particles. To validate our code, called Pic-Vert, we consider a 3d electrostatic Landau-damping simulation as well as a 2d3v transverse instability of magnetized electron holes. Performance results on a 24-core Intel Skylake hardware confirm the effectiveness of our algorithm, in particular its high throughput and its ability to cope with fast moving particles. A paper describing this work was published at Euro-par [13] and is described in more details in Yann Barsamian's PhD thesis [6].

7.6. 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 recently 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 [22], 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 have been accepted for publication at PPoPP'19.

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 published at PLDI'18 [12].

7.7. Program verification and formal languages

Participant: Arthur Charguéraud.

- 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 describing this work was published at ESOP'18 [15].
- A. Charguéraud, together with Ralf Jung and Jan-Oliver Kaiser and Derek Dreyer (MPI-SWS), Robbert Krebbers (Delft University of Technology), Jacques-Henri Jourdan (Inria), Joseph Tassarotti (Carnegie Mellon University), and Amin Timany (KU Leuven), developed MoSel, a general and extensible Coq framework for carrying out separation-logic proofs mechanically using an interactive proof assistant. This tool extends the Iris Proof Mode (IPM) to make it applicable to both affine and linear separation logics (and combinations thereof), and to provide generic tactics that can be easily extended to account for the bespoke connectives of the logics with which it is instantiated. To demonstrate the effectiveness of MoSeL, the tool has been instantiated to provide effective tactical support for interactive and semi-automated proofs in six very different separation logics. This work was published at ICFP'18 [17].
- A. Charguéraud advised Ramon Fernandez for a 4-month internship. The aim of that internship was to formalize, using the Coq proof assistant, several data layout transformations such as the transformation from an array of structures to a structure of arrays (AoS-to-SoA). Such transformations are routinely employed to develop high-performance code. Ramon investigated the literature on data layout transformations, listed the most useful transformations exploited in practice, and identified several core transformations from which almost all others can be derived. He then successfully carried out proofs of semantic preservation for the three most important transformations: field grouping, tiling, and AoS-to-SoA.
- 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* [24]. 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 appeared at the international conference Web Programming [14].

7.8. Flexible Runtime System with High Throughput for Many-to-Many Data Stream Problems

Participants: Paul Godard, Vincent Loechner, Cédric Bastoul.

In the context of our collaboration with the Caldera company, we are interested in high throughput data stream problems, that require low latency, maximal bandwidth usage, and that avoid starvations. We suppose that we receive jobs from an external system through a queue, each job including a description of its computation needs, output data requirements and output locations.

The computations are distributed on a cluster organized in a many-to-many logical topology where one or many computing tasks (producers) send data to one or many consumer tasks (consumers). The runtime system is orchestrated by a centralized scheduler, which decomposes jobs into tasks and dynamically assigns them to producers. The producers perform the computations and send their output data to the consumers. The consumers collect and order output data to make them available to the final user.

We implemented our framework, and performed some experiments on a real-world use case: real time professional digital printing, that may require tens of Gbit/s sustained output rates. We show in our measurements that our system scales and reaches data rates that are close to the maximum throughput of our experimental hardware. The architecture as a cluster and using the standard TCP/IP network protocol allow our system to be highly adaptive to the user's requirements. We are in the process of writing a paper describing our framework architecture for many-to-many data stream problems and results.

7.9. Visual Program Manipulation in the Polyhedral Model

Participants: Cédric Bastoul, Oleksandr Zinenko, Stéphane Huot.

While a plethora of libraries and frameworks focus on expressing parallelism, identifying and extracting it remains a challenging task. Automatic parallelization relies on imprecise heuristics resulting in cumbersome manual code analysis and transformation in case of underperformance. Alternatively, directive-based approaches often require transforming the program from scratch when a slightly modified version of an automatically-computed transformation would suffice. We propose an interactive visual approach building on the polyhedral model that (1) visualizes exact dependences and parallelism, (2) decomposes a complex automatically-computed transformation into simple steps for replay and easier modification, and (3) allows for directly manipulating the visual representation as a means of transforming the program with immediate feedback. User studies suggest that our visualization is understood by experts and non-experts alike, and that it may favor an exploratory approach to transformation. Finally, an eye-tracking study suggests that programmers may resort to visualizations instead of code if visualizations are clearly efficient for a given task.

This is a joint work with PARKAS team at Inria Paris (contact: Oleksandr Zinenko) and MJOLNIR team at Inria Lille (contact: Stéphane Huot), published in TACO [10].

7.10. A language extension set to generate adaptive versions automatically

Participants: Maxime Schmitt, Cédric Bastoul.

A large part of the development effort of compute-intensive applications is devoted to optimization, i.e., achieving the computation within a finite budget of time, space or energy. Given the complexity of modern architectures, writing simulation applications is often a two-step workflow. Firstly, developers design a sequential program for algorithmic tuning and debugging purposes. Secondly, experts optimize and exploit possible approximations of the original program to scale to the actual problem size. This second step is a tedious, time-consuming and error-prone task. During this year, we investigated language extensions and compiler tools to achieve that task semi-automatically in the context of approximate computing. We identified the semantic and syntactic information necessary for a compiler to automatically handle approximation and adaptive techniques for a particular class of programs. We proposed a set of language extensions generic enough to provide the compiler with the useful semantic information when approximation is beneficial. We implemented the compiler infrastructure to exploit these extensions and to automatically generate the adaptively approximated version of a program. We conducted an experimental study of the impact and expressiveness of our language extension set on various applications.

These language extensions and the underlying compiler infrastructure are a significant output of collaboration with Inria Nancy - Grand Est team TONUS, specialized on applied mathematics (contact: Philippe Helluy), to bring models and techniques from this field to compilers. A paper presenting these extensions has been accepted to the OGST journal, targeting typical end-users.

CASH Team

7. New Results

7.1. Monoparametric Tiling of Polyhedral Programs

Participant: Christophe Alias.

Tiling is a crucial program transformation, adjusting the ops-to-bytes balance of codes to improve locality. Like parallelism, it can be applied at multiple levels. Allowing tile sizes to be symbolic parameters at compile time has many benefits, including efficient autotuning, and run-time adaptability to system variations. For polyhedral programs, parametric tiling in its full generality is known to be non-linear, breaking the mathematical closure properties of the polyhedral model. Most compilation tools therefore either perform fixed size tiling, or apply parametric tiling in only the final, code generation step.

We introduce monoparametric tiling, a restricted parametric tiling transformation. We show that, despite being parametric, it retains the closure properties of the polyhedral model. We first prove that applying monoparametric partitioning (i) to a polyhedron yields a union of polyhedra, and (ii) to an affine function produces a piecewise-affine function. We then use these properties to show how to tile an entire polyhedral program. Our monoparametric tiling is general enough to handle tiles with arbitrary tile shapes that can tesselate the iteration space (e.g., hexagonal, trapezoidal, etc). This enables a wide range of polyhedral analyses and transformations to be applied.

This is a joint work with Guillaume Iooss (Inria Parkas) and Sanjay Rajopadhye (Colorado State University).

This work is under submission [12].

7.2. Improving Communication Patterns in Polyhedral Process Networks

Participant: Christophe Alias.

Process networks are a natural intermediate representation for HLS and more generally automatic parallelization. Compiler optimizations for parallelism and data locality restructure deeply the execution order of the processes, hence the read/write patterns in communication channels. This breaks most FIFO channels, which have to be implemented with addressable buffers. Expensive hardware is required to enforce synchronizations, which often results in dramatic performance loss. In this paper, we present an algorithm to partition the communications so that most FIFO channels can be recovered after a loop tiling, a key optimization for parallelism and data locality. Experimental results show a drastic improvement of FIFO detection for regular kernels at the cost of a few additional storage. As a bonus, the storage can even be reduced in some cases.

This work has been published in the HIP3ES workshop [1]

7.3. FIFO Recovery by Depth-Partitioning is Complete on Data-aware Process Networks

Participant: Christophe Alias.

In this paper, we build on our algorithm for FIFO recovery based on depth partitioning. We describe a class of process networks where the algorithm can recover all the FIFO channels. We point out the limitations of the algorithm outside of that class. Experimental results confirm the completeness of the algorithm on the class and reveal good performance outside of the class.

This work is under submission [9]

7.4. Parallel code generation of synchronous programs for a many-core architecture

Participant: Matthieu Moy.

Embedded systems tend to require more and more computational power. Many-core architectures are good candidates since they offer power and are considered more time predictable than classical multi-cores. Data-flow Synchronous languages such as Lustre or Scade are widely used for avionic critical software. Programs are described by networks of computational nodes. Implementation of such programs on a many-core architecture must ensure a bounded response time and preserve the functional behavior by taking interference into account. We consider the top-level node of a Lustre application as a software architecture description where each sub-node corresponds to a potential parallel task. Given a mapping (tasks to cores), we automatically generate code suitable for the targeted many-core architecture. This minimizes memory interferences and allows usage of a framework to compute the Worst-Case Response Time.

This is a joint work with Amaury Graillat, Pascal Raymond (IMAG) and Benoît Dupont de Dinechin (Kalray).

This work has been published at the DATE conference [6].

7.5. Estimation of the Impact of Architectural and Software Design Choices on Dynamic Allocation of Heterogeneous Memories

Participant: Matthieu Moy.

Reducing energy consumption is a key challenge to the realization of the Internet of Things. While emerging memory technologies may offer power reduction, they come with major drawbacks such as high latency or limited endurance. As a result, system designers tend to juxtapose several memory technologies on the same chip. This paper studies the interactions between dynamic memory allocation and architectural choices regarding this heterogeneity. We provide cycle accurate simulations of embedded platforms with various memory technologies and we show that different dynamic allocation strategies have a major impact on performance. We demonstrate that interesting performance gains can be achieved even for a low fraction of heap objects in fast memory, but only with a clever data placement strategy between memory banks.

This is a joint work with Tristan Delizy, Kevin Marquet, Tanguy Risset, Guillaume Salagnac (Inria Socrate) and Stéphane Gros (eVaderis).

This work has been published at the French Compas workshop [8] and the RSP Symposium [2].

7.6. Dataflow-explicit futures

Participant: Ludovic Henrio.

A future is a place-holder for a value being computed, and we generally say that a future is resolved when the associated value is computed. In existing languages futures are either implicit, if there is no syntactic or typing distinction between futures and non-future values, or explicit when futures are typed by a parametric type and dedicated functions exist for manipulating futures. We defined a new form of future, named data-flow explicit futures [38], with specific typing rules that do not use classical parametric types. The new futures allow at the same time code reuse and the possibility for recursive functions to return futures like with implicit futures, and let the programmer declare which values are futures and where synchronisation occurs, like with explicit futures. We prove that the obtained programming model is as expressive as implicit futures but exhibits a different behaviour compared to explicit futures. The current status of this work is the following:

- A paper showing formally the difference between implicit and explicit futures is under submission
- We are working with collaborators from University of Uppsala and University of Oslo on the design of programming constructs mixing implicit and dataflow-explicit futures
- Amaury Maillé will do his internship in the Cash team (advised by Matthieu Moy and Ludovic Henrio), working on an implementation of dataflow-explicit futures and further experiments with the model.

7.7. Locally abstract globally concrete semantics

Participant: Ludovic Henrio.

This research direction aims at designing a new way to write semantics for concurrent languages. The objective is to design semantics in a compositional way, where each primitive has a local behavior, and to adopt a style much closer to verification frameworks so that the design of an automatic verifier for the language is easier. The local semantics is expressed in a symbolic and abstract way, a global semantics gathers the abstract local traces and concretizes them. We have a reliable basis for the semantics of a simple language (a concurrent while language) and for a complex one (ABS), but the exact semantics and the methodology for writing it is still under development, we expect to submit a journal article during 2019 on the subject.

This is a joint with Reiner Hähnle (TU Darmstadt), Einar Broch Johnsen, Crystal Chang Din, Lizeth Tapia Tarifa (Univ Oslo), Ka I Pun (Univ Oslo and Univ of applied science).

7.8. Memory consistency for heterogeneous systems

Participant: Ludovic Henrio.

Together with Christoph Kessler (Linköping University), we worked on the formalization of the cache coherency mechanism used in the VectorPU library developed at Linköping University. Running a program on disjoint memory spaces requires to address memory consistency issues and to perform transfers so that the program always accesses the right data. Several approaches exist to ensure the consistency of the memory accessed, we are interested here in the verification of a declarative approach where each component of a computation is annotated with an access mode declaring which part of the memory is read or written by the component. The programming framework uses the component annotations to guarantee the validity of the memory accesses. This is the mechanism used in VectorPU, a C++ library for programming CPU-GPU heterogeneous systems and this article proves the correctness of the software cache-coherence mechanism used in the library. Beyond the scope of VectorPU, this article can be considered as a simple and effective formalisation of memory consistency mechanisms based on the explicit declaration of the effect of each component on each memory space. This year, we have the following new results:

- provided a formalization showing the correctness of VectorPU approach (published in 4PAD 2018, a symposium affiliated to HPCS).
- extended the work to support the manipulation of overlapping array (submitted as an extended version of the 4PAD paper)

We now plan to extend the work with support for concurrency.

7.9. PNets: Parametrized networks of automata

Participant: Ludovic Henrio.

pNets (parameterised networks of synchronised automata) are semantic objects for defining the semantics of composition operators and parallel systems. We have used pNets for the behavioral specification and verification of distributed components, and proved that open pNets (i.e. pNets with holes) were a good formalism to reason on operators and parameterized systems. This year, we have the following new results:

- A weak bisimulation theory for open pNets is under development (a strong isimulation had already been defined in the past) and its properties are being proven, especially in terms of compositionality. This work is realized with Eric Madelaine (Inria Sophia-Antipolis) and Rabéa Ameur Boulifa (Telecom ParisTech).
- A translation from BIP model to open pNets is being formalized and encoded, this work is done in collaboration with Simon Bliudze (Inria Lille).

These works are under progress and should be continued in 2019.

7.10. Decidability results on the verification of phaser programs

Participant: Ludovic Henrio.

Together with Ahmed Rezine and Zeinab Ganjei (Linköping University) we investigated the possibility to analyze programs with phasers (a construct for synchronizing processes that generalizes locks, barrier, and publish-subscribe patterns). They work with signal and wait messages from the processes (comparing the number of wait and signal received to synchronize the processes). We proved that in many conditions, if the number of phasers or processes cannot be bounded, or if the difference between the number of signal and the number of wait signal is unbounded, then many reachability problems are undecidable. We also proposed fragments where these problems become decidable, and proposed an analysis algorithm in these cases. The results are currently under review in a conference.

7.11. Practicing Domain-Specific Languages: From Code to Models

Participant: Laure Gonnord.

Together with Sebastien Mosser, we proposed a new Domain- Specific Language course at the graduate level whose objectives is to reconciliate concepts coming from Language Design as well as Modeling domains. We illustrate the course using the reactive systems application domain, which prevents us to fall back in a toy example pitfall. This paper describes the nine stages used to guide students through a journey starting at low-level C code to end with the usage of a language design workbench. This course was given as a graduate course available at Université Côte d'Azur (8 weeks, engineering-oriented) and École Normale Supérieure de Lyon (13 weeks, research-oriented).

The results have been published in a national software engineering conference [4] and the Models Educator Symposium [5].

7.12. Polyhedral Dataflow Programming: a Case Study

Participant: Laure Gonnord.

With Lionel Morel and Romain Fontaine (Insa Lyon), we have studied the benefits of jointly using polyhedral compilation with dataflow languages. We have proposed to expend the parallelization of dataflow programs by taking into account the parallelism exposed by loop nests describing the internal behavior of the program's agents. This approach is validated through the development of a prototype toolchain based on an extended version of the SigmaC language. We demonstrated the benefit of this approach and the potentiality of further improvements on several case studies.

The results have been published in the Sbac-PAD conference on High Performance computing [3].

7.13. Semantic Array Dataflow Analysis

Participants: Laure Gonnord, Paul Iannetta.

Together with Lionel Morel (Insa/CEA) and Tomofumi Yuki (Inria, Rennes), we revisited the polyhedral model's key analysis, dependency analysis. The semantic formulation we propose allows a new definition of the notion of dependency and the computation of the dependency set. As a side effect, we propose a general algorithm to compute *an over-approximation of* the dependency set of general imperative programs.

We argue that this new formalization will later allow for a new vision of the polyhedral model in terms of semantics, which will help us fully characterize its expressivity and applicability. We also believe that abstract semantics will be the key for designing an approximate abstract model in order to enhance the applicability of the polyhedral model.

The results is published in a research report [11].

7.14. Static Analysis Of Binary Code With Memory Indirections Using Polyhedra

Participant: Laure Gonnord.

Together with Clement Ballabriga, Julien Forget, Giuseppe Lipari, and Jordy Ruiz (University of Lille), we proposed a new abstract domain for static analysis of binary code. Our motivation stems from the need to improve the precision of the estimation of the Worst-Case Execution Time (WCET) of safety-critical real-time code. WCET estimation requires computing information such as upper bounds on the number of loop iterations, unfeasible execution paths, etc. These estimations are usually performed on binary code, mainly to avoid making assumptions on how the compiler works. Our abstract domain, based on polyhedra and on two mapping functions that associate polyhedra variables with registers and memory, targets the precise computation of such information. We prove the correctness of the method, and demonstrate its effectiveness on benchmarks and examples from typical embedded code.

The results have been accepted to VMCAI'19 on Model Checking and Abstract Interpretation [7].

7.15. Polyhedral Value Analysis as Fast Abstract Interpretation

Participant: Laure Gonnord.

Together with Tobias Grosser, (ETH Zurich, Switzerland), Siddhart Bhat, (IIIT Hydrabad, India), Marcin Copik (ETH Zurich, Switzerland), Sven Verdoolaege (Polly Labs, Belgium) and Torsten Hoefler (ETH Zurich, Switzerland), we tried to bridge the gap between the well founded classical abstract interpretation techniques and their usage in production compilers.

We formulate the polyhedral value analysis (a classical algorithm in production compilers like LLVM, scalar evolution based on Presburger set as abstract interpretation), present a set of fast join operators, and show that aggressively falling back to top (rather than continuing with approximations) results in a scalable analysis. By formally describing the required analysis, we provide the necessary theoretical foundations for analysing large program systems with hundred thousands of loops and complex control flow structures at a precision high enough to cater for high-precision users such as polyhedral optimization frameworks, at a compile-time cost comparable with just compiling the application.

The paper is under redaction process.

CORSE Project-Team

7. New Results

7.1. Profiling Feedback based Optimizations and Performance Debugging

Participants: Fabrice Rastello, Diogo Sampaio, Fabian Gruber, Christophe Guillon [STMicroelectronics], Antoine Moynault [STMicroelectronics], Changwan Hong [OSU, USA], Aravind Sukumaran-Rajam [OSU, USA], Jinsung Kim [OSU, USA], Prashant Singh Rawat [OSU, USA], Sriram Krishnamoorthy [PNNL, USA], Louis-Noël Pouchet [CSU, USA], P. Sadayappan [OSU, USA].

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. Our contributions in this area are twofold: (1) we developed a new technique that combines abstract simulation and sensitive analysis that allows to pinpoint performance bottleneck; (2) we developed a new technique to build a polyhedral representation out of an execution trace that allows to provide feedback on possible missed transformations.

7.1.1. Compiler Optimization for GPUs Using Bottleneck Analysis

Optimizing compilers generally use highly simplified performance models due to the significant challenges in developing accurate analytical performance models for complex computer systems. In this work, we develop an alternate approach to performance modeling using abstract execution of GPU kernel binaries. We use the performance model to predict the bottleneck resource for a given kernel's execution through differential analysis by performing multiple abstract executions with varying machine parameters. The bottleneck analysis is then used to develop an automated search through a configuration space of different grid reshaping, thread/block coarsening, and loop unrolling factors. Experimental results using a number of benchmarks from the Parboil/Rodinia/SHOC suites demonstrate the effectiveness of the approach. The bottleneck analysis is also shown to be useful in assisting high-level domain-specific code generators for GPUs.

This work is the fruit of the collaboration 9.4.1.1 with OSU. It has been presented at the ACM/SIGPLAN conference on Programming Language Design and Implementation, PLDI 2018.

7.1.2. Data-Flow/Dependence Profiling for Structured Transformations

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, 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 rescheduling. Our tool can handle both inter- and intraprocedural aspects of the program in a unified way, thus providing structured interprocedural transformation feedback.

This work is the fruit of the collaboration 9.4.1.1 with CSU and the past collaboration Nano2017 with STMicroelectronics. It has been submitted for presentation at the ACM conference on Principles and Practice of Parallel Programming, PPoPP 2019.

7.2. Combined Scheduling and Register Allocation

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

Register allocation is one of the most studied compiler optimization but its impact on performance is highly coupled with scheduling. Recent advances on computer simulation and artificial intelligence lead to application kernels with very high register pressure. Our contributions in this area consist in developing new scheduling schemes that both expose SIMD parallelism and register reuse.

7.2.1. Register Optimizations for Stencils on GPUs

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 work, 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 9.4.1.1 with OSU. It has been presented at the ACM/SIGPLAN Symposium on Principles and Practice of Parallel Programming, PPoPP 2018.

7.2.2. Associative instruction reordering to alleviate register pressure

Register allocation is generally considered a practically solved problem. For most applications, the register allocation strategies in production compilers are very effective in controlling the number of loads/stores and register spills. However, existing register allocation strategies are not effective and result in excessive register spilling for computation patterns with a high degree of many-to-many data reuse, e.g., high-order stencils and tensor contractions. We develop a source-to-source instruction reordering strategy that exploits the flexibility of reordering associative operations to alleviate register pressure. The developed transformation module implements an adaptable strategy that can appropriately control the degree of instruction-level parallelism, while relieving register pressure. The effectiveness of the approach is demonstrated through experimental results using multiple production compilers (GCC, Clang/LLVM) and target platforms (Intel Xeon Phi, and Intel x86 multi-core).

This work is the fruit of the collaboration 9.4.1.1 with OSU. It has been presented at ACM/IEEE International Conference for High Performance Computing, Networking, Storage, and Analysis, SC 2018.

7.3. Runtime Verification and Monitoring

Participants: Raphael Jakse, Yliès Falcone, Jean Francois Mehaut, Srdan Krstic, Giles Reger, Dmitriy Traytel, Hosein Nazarpour, Mohamad Jaber, Marius Bozga, Saddek Bensalem, Salwa Kobeissi, Adnan Utayim.

We report on several contributions related with the runtime verification and monitoring of systems. We address several aspects such as the instrumentation, the understanding and classification of existing concepts and tools, the definition of frameworks for monitoring distributed systems and a case study on monitoring smart homes.

7.3.1. Interactive Runtime Verification: Formal Models, Algorithms, and Implementation

Interactive runtime verification (i-RV) combines runtime verification and interactive debugging. Runtime verification consists in studying a system at runtime, looking for input and output events to discover, check or enforce behavioral properties. Interactive debugging consists in studying a system at runtime in order to discover and understand its bugs and fix them, inspecting interactively its internal state. 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 runtime verification.

In this work, we provide a formal model for a program execution under a debugger, which we compose with a general model of a monitor and a scenario to model the interactively verified program. We provide guarantees on the verdicts issued by the monitor using the instrumentation provided by the debugger. We provide an algorithmic view of this model suitable for producing implementations, and we present Verde, an implementation based on GDB to interactively verify C programs. We built a set of experiments using Verde to assess usefulness of Interactive Runtime Verification and performance of our implementation. Our results show that though debugger-based instrumentation incurs non-trivial performance costs, i-RV is appliable performance-wise in a variety of cases and helps studying bugs.

This work has been submitted at the ACM Transactions on Software Engineering and Methodology (TOSEM).

7.3.2. A Taxonomy for Classifying Runtime Verification Tools

Over the last 15 years Runtime Verification (RV) has grown into a diverse and active field, which has stimulated the development of numerous theoretical frameworks and tools. Many of the tools are at first sight very different and challenging to compare. Yet, there are similarities. In this work, we classify RV tools within a high-level taxonomy of concepts. We first present this taxonomy and discuss the different dimensions. Then, we survey RV tools and classify them according to the taxonomy. This work constitutes a snapshot of the current state of the art and enables a comparison of existing tools.

This work has been published in the proceedings of the 18th International Conference on Runtime Verification.

7.3.3. Bringing Runtime Verification Home

We use runtime verification (RV) to check various specifications in a smart apartment. The specifications can be broken down into three types: be- havioral correctness of the apartment sensors, detection of specific user activities (known as activities of daily living), and composition of specifications of the previous types. The context of the smart apartment provides us with a complex system with a large number of components with two different hierarchies to group specifications and sensors: geographically within the same room, floor or globally in the apartment, and logically following the different types of specifications. We leverage a recent approach to decentralized RV of decentralized specifications, where monitors have their own specifications and communicate together to verify more general specifications. This allows us to re-use specifications, and combine them to: (1) scale beyond existing centralized RV techniques, and (2) greatly reduce computation and communication costs.

This work has been published in the proceedings of the 18th International Conference on Runtime Verification.

7.3.4. Tracing Distributed Component-Based Systems, a Brief Overview

We overview a framework for tracing asynchronous distributed component-based systems with multiparty interactions managed by distributed schedulers. Neither the global state nor the total ordering of the system events is available at runtime. We instrument the system to retrieve local events from the local traces of the schedulers. Local events are sent to a global observer which reconstructs on-the-fly the global traces that are compatible with the local traces, in a concurrency-preserving and communication-delay insensitive fashion. The global traces are represented as an original lattice over partial states, such that any path of the lattice projected on a scheduler represents the corresponding lo- cal partial trace according to that scheduler (soundness), and all possible global traces of the system are recorded (completeness).

This work has been published in the proceedings of the 18th International Conference on Runtime Verification.

7.3.5. Can We Monitor All Multithreaded Programs?

Runtime Verification(RV)is a lightweight formal method which consists in verifying that an execution of a program is correct wrt a specification. The specification formalizes with properties the expected correct behavior of the system. Programs are instrumented to extract necessary information from the execution and feed it to monitors tasked with checking the properties. From the perspective of a monitor, the system is a black box; the trace is the only system information provided. Parallel programs generally introduce an added level of complexity on the program execution due to concurrency. A concurrent execution of a parallel program is best represented as a partial order. A large number of RV approaches generate monitors using formalisms that rely on total order, while more recent approaches utilize formalisms that consider multiple traces.

We made a tutorial where we review some of the main RV approaches and tools that handle multithreaded Java programs. We discuss their assumptions, limitations, expressiveness, and suitability when tackling parallel programs such as producer- consumer and readers-writers. By analyzing the interplay between specification formalisms and concurrent executions of programs, we identify four questions RV practitioners may ask themselves to classify and determine the situations in which it is sound to use the existing tools and approaches.

This work has been published in the proceedings of the 18th International Conference on Runtime Verification.

7.3.6. Facilitating the Implementation of Distributed Systems with Heterogeneous Interactions

We introduce HDBIP an extension of the Behavior Interaction Priority (BIP) framework. BIP is a component-based framework with a rigorous operational semantics and high-level and expressive interaction model. HDBIP extends BIP interaction model by allowing heterogeneous interactions targeting distributed systems. HDBIP allows both multiparty and direct send/receive interactions that can be directly mapped to an underlying communication library. Then, we present a correct and efficient code generation from HDBIP to C++ implementation using Message Passing Interface (MPI). We present a non-trivial case study showing the effectiveness of HDBIP.

This work has been published in the proceedings of the 14th International Conference on integrated Formal Methods.

7.3.7. Modularizing Behavioral and Architectural Crosscutting Concerns in Formal Component-Based Systems

We define a method to modularize crosscutting concerns in Component-Based Systems (CBSs) expressed using the Behavior Interaction Priority (BIP) framework. Our method is inspired from the Aspect Oriented Programming (AOP) paradigm which was initially conceived to support the separation of concerns during the development of monolithic systems. BIP has a formal operational semantics and makes a clear separation between architecture and behavior to allow for compositional and incremental design and analysis of systems. We distinguish local from global aspects. Local aspects model concerns at the component level and are used to refine the behavior of components. Global aspects model concerns at the architecture level, and hence refine communications (synchronization and data transfer) between components. We formalize local and global aspects as well as their composition and integration into a BIP system through rigorous transformation primitives. We present AOP-BIP, a tool for Aspect-Oriented Programming of BIP systems, demonstrate its use to modularize logging, security, and fault tolerance in a network protocol, and discuss its possible use in runtime verification of CBSs.

This work has been published in the Journal of Logical and Algebraic Methods in Programming.

7.4. Numa MeMory Analyzer

Participants: François Trahay [Télécom SudParis], Manuel Selva, Lionel Morel [CEA], Kevin Marquet [INSA Lyon].

Non Uniform Memory Access (NUMA) architectures are nowadays common for running High-Performance Computing (HPC) applications. In such architectures, several distinct physical memories are assembled to create a single shared memory. Nevertheless, because there are several physical memories, access times to these memories are not uniform depending on the location of the core performing the memory request and on the location of the target memory. Hence, threads and data placement are crucial to efficiently exploit such architectures. To help in taking decision about this placement, profiling tools are needed. NUMA MeMory Analyzer (NumaMMA) is a new profiling tool for understanding the memory access patterns of HPC applications. NumaMMA combines efficient collection of memory traces using hardware mechanisms with original visualization means allowing to see how memory access patterns evolve over time. The information reported by NumaMMA allows to understand the nature of these access patterns inside each object allocated by the application. We show how NumaMMA can help understanding the memory patterns of several HPC applications in order to optimize them and get speedups up to 28% over the standard non optimized version.

This work has been published in the 47th International Conference on Parallel Processing - ICPP 2018.

7.5. Towards an Easier Way to Program FPGAs in an HPC Context

Participants: Georgios Christodoulis, Manuel Selva, Francois Broquedis, Frederic Desprez, Olivier Muller [TIMA].

Heterogeneity in HPC nodes appears as a promising solution to improve the execution of a wide range of scientific applications, regarding both performance and energy consumption. Unlike CPUs and GPUs, FPGAs can be configured to fit the application needs, making them an appealing target to extend traditional heterogeneous HPC architectures. However, exploiting them requires an in-depth knowledge of low-level hardware and high expertise on vendor-provided tools, which should not be the primary concern of HPC application programmers. In the context of the Persyval HEAVEN project, we proposed a framework enabling a more straightforward development of scientific applications over FPGA enhanced platforms. Our solution requires the minimum knowledge of the underlying architecture, as well as fewer changes to the existing code. To fulfill these requirements, we extended the StarPU task programming library that initially targets heterogeneous architectures to support FPGAs. We used Vivado HLS, a high-level synthesis tool to deliver efficient hardware implementations of the tasks from high-level languages like C/C++. Our solution, validated on a blocking version of the matrix multiplication algorithm, offers an easier way to exploit FPGAs from an HPC application. We also conducted some preliminary experiments to validate our proof-of-concept implementation regarding performance.

This work has been published in the 13th International Symposium on Reconfigurable Communication-centric Systems-on-Chip and obtained the best paper award.

7.6. Automatic IPC Profile Analysis to Detect Phases in HPC Application

Participants: Mathieu Stoffel, François Broquedis, Frederic Desprez, Abdelhafid Mazouz [Atos/Bull], Philippe Rols [Atos/Bull].

Mathieu Stoffel started his PhD in February 2018 on a CIFRE contract with Atos/Bull. The purpose of this work is to enhance the energy consumption of HPC applications on large-scale platforms. The first phase of the thesis project consists in an in-depth study of the evolution of the metrics characterizing the state of the supercomputer during the execution of a highly parallel application. Indeed, the utilization rates of the different components of the HPC system may demonstrate extreme variations during the execution of the aforementioned application. These variations are sometimes subject to repeat themselves on a regular basis during the application execution. We refer to this phenomena as application "phases". In this context, we already generated precise IPC profiles out of many benchmarks and real-life applications and we worked on a methodology to adapt the CPU frequency based on these profiles. This part of the thesis has been published in an IEEE Cluster workshop (HPCMASPA). Currently, we are working on a detection tool for the application phases. It will implement an automated reconfiguration of the parameters of the HPC system during the execution of the application, in relation with the type of phase being executed. By doing so, the tool will aim at optimizing the energy consumption associated with the execution of the application, by adapting the state of the HPC systems all along the aforesaid execution.

7.7. Teaching of Algorithms, Programming, and Debugging

Participants: Florent Bouchez-Tichadou, Theo Barollet, Aurelien Flori, Thomas Herve.

7.7.1. Teaching Algorithms using Problem and Challenge Based Learning

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 has been published in the 23rd International Conference on Innovation and Technology in Computer Science Education, ITiCSE 2018.

7.7.2. Data Structures Visualization at Runtime

Debuggers are powerful tools to observe a program behaviour and find bugs but they are not often used by developers and especially beginners because of the hard learning curve of such tools. They provide information on low level data but are not able to analyze higher level elements such as data structures. This work tries to provide a more intuitive representation of the program execution to ease debugging and algorithms understanding.

We have a basic prototype, Moly, which is a GDB extension (GNU Project Debugger) to explore a program runtime memory and analyze its data structures. It also provides an interface with an external visualizer, Lotos, through a formatted output. Running Moly along with a dedicated visualizer should allow a programmer to spot bugs easier by seeing the subsequent whole memory states of the program and some data structures information.

The current status of Moly allows a programmer to explore all attainable memory at any point during the debug process, and already provides minimum information about the possible properties of the data structures, such as recognizing graphs, trees, or linked lists. Future work includes recognizing access patterns to the structures to extract for instance visit patterns and higher-level properties (such as the breaking of data structure properties between break-points).

The external visualizer, Lotos, is still in its early stages of development and was enough to make a proof-of-concept that it is possible to display via a web browser the information gathered by Moly. Our plans is to redesign this part from scratch using the knowledge gaining during the writing of this prototype.

7.7.3. AppoLab: an Online Platform to Engage Students in Their Learning

Classical teaching of algorithms and low-level data structures at the L2 european level is often tedious and unappealing to students, with much of the time being spent on analysing and devising algorithms for textbook cases, such as sorting lists of integers, visiting linked lists or trees, etc.

Using Problem-Based Learning helps to alleviate this problem, by presenting more complex problems to handle, hence engaging more students in their learning. This work revolves around the design of a learning platform that includes gamification in PBL. AppoLab is in its core a server that has scripted "exercices". Students can communicate with the server either manually, using telnet; but ultimately, they will need to script the communication also from their side, since the server will gradually impose constraints on the problems such as timeouts or large input sizes.

This preliminary work was used this year in some parts of an Algorithm course at the L2 level, and has received positive feedback from the students. This encourages us to continue this development and study more precisely the impact it has on students' engagement in their learning.

PACAP Project-Team

7. New Results

7.1. Compilation and Optimization

Participants: Arif Ali Ana-Pparakkal, Loïc Besnard, Rabab Bouziane, Sylvain Collange, Byron Hawkins, Imane Lasri, Kévin Le Bon, Erven Rohou.

7.1.1. Optimization in the Presence of NVRAM

Participants: Rabab Bouziane, Erven Rohou, Bahram Yarahmadi.

Energy-efficiency has become one major challenge in both embedded and high-performance computing. Different approaches have been investigated to solve the challenge, e.g., heterogeneous multicore, system runtime and device-level power management, or deployment of emerging non-volatile memories (NVMs), such as Spin-Transfer Torque RAM (STT-RAM), which inherently have quasi-null leakage. This enables to reduce the static power consumption, which tends to become dominant in modern systems. The usage of NVM in memory hierarchy comes however at the cost of expensive write operations in terms of latency and energy.

7.1.1.1. Silent-Stores

We propose [31] 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.

7.1.1.2. Variable Retention Time

In order to mitigate expensive writes, we leverage the notion of *δ-worst-case execution time* (δ-WCET), which consists of partial WCET estimates [32]. From program analysis, δ-WCETs are determined and used to safely allocate data to NVM memory banks with variable data retention times. The δ-WCET analysis computes the WCET between any two locations in a function code, i.e., between basic blocks or instructions. Our approach is validated on the Mälardalen benchmark suite and significant memory dynamic energy reductions (up to 80 %, and 66 % on average) are reported.

This research is done in collaboration with Abdoulaye Gamatié at LIRMM (Montpellier) within the context the ANR project CONTINUUM. Results are detailed in the PhD thesis document of Rabab Bouziane, defended in December 2018 [20].

7.1.1.3. Efficient checkpointing for intermittently-powered systems

Future internet of things (IoT) ultra low-power micro controllers do not have any battery. Instead they harvest energy from the environment such as solar or radio and store it to a capacitor. However, one of the unique problems with these energy harvesting devices is the unstable energy supply which causes frequent power failures during the execution of the program. As a result, a program may not be able to terminate with one power cycle. A solution to this problem consists in using non-volatile memory (NVM) such as FLASH or Ferroelectric RAM (FRAM) and checkpointing the volatile state of the program to the non-volatile memory regularly. The program can resume its execution when the power is back. However, checkpointing regularly at runtime has overhead, and poses some memory inconsistency problems [47]. By statically analyzing the program binary code, we propose to insert checkpoints in the proper places in the program to decrease the amount of checkpointing overhead at runtime. Also, a compiler can give hints to runtime system about whether to do the checkpoint or not. Concerning the reduction of checkpointing overhead, we are analyzing the binary code of the program for estimating the energy of each section of the program. We rely on Heptane, which is originally designed for estimating worst-case execution time. However, by giving energy cost to each ISA instruction, we can estimate the energy consumption of the sections of the program for processors like MSP430 or Arm-cortex m0+ which are typical low-power embedded processors. With this information, we insert checkpoints into the LLVM IR and also apply optimizations in order to have better performance and energy efficiency at runtime.

This research is done within the context of the project IPL ZEP.

7.1.2. Dynamic Binary Optimization

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

Modern hardware features can boost the performance of an application, but software vendors are often limited to the lowest common denominator to maintain compatibility with the spectrum of processors used by their clients. Given more detailed information about the hardware features, a compiler can generate more efficient code, but even if the exact CPU model is known, manufacturer confidentiality policies leave substantial uncertainty about precise performance characteristics. In addition, the activity of other programs colocated in the same runtime environment can have a dramatic effect on application performance. For example, if a shared CPU cache is being heavily used by other programs, memory access latencies may be orders of magnitude longer than those recorded during an isolated profiling session, and instruction scheduling based on such profiles may lose its anticipated advantages. Program input can also drastically change the efficiency of statically compiled code, yet in many cases is subject to total uncertainty until the moment the input arrives during program execution. We have developed FITTCHOOSER [30] to defer optimization of a program's most processor-intensive functions until execution time. FITTCHOOSER begins by profiling the application to determine the performance characteristics that are in effect for the present execution, then generates a set of candidate variations and dynamically links them in succession to empirically measure which of them performs best. The underlying binary instrumentation framework Padrone allows for selective transformation of the program without otherwise modifying its structure or interfering with the flow of execution, making it possible for FITTCHOOSER to minimize the overhead of its dynamic optimization process. Our experimental evaluation demonstrates up to 19 % speedup on a selection of programs from the SPEC CPU 2006 and PolyBench suites while introducing less than 1 % overhead. The FITTCHOOSER prototype achieves these gains with a minimal repertoire of optimization techniques taken from the static compiler itself, which not only testifies to the effectiveness of dynamic optimization, but also suggests that further gains can be achieved by expanding FITTCHOOSER's repertoire of program transformations to include more diverse and more advanced techniques.

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

Nowadays almost every device has parallel architecture, hence parallelization is almost always desirable. However, parallelizing legacy running programs is very challenging. That is due to the fact that usually source code is not available, and runtime parallelization is challenging. Also, detecting parallelizable code is difficult, due to possible dependencies and different execution paths that are undecidable statically. Therefore, speculation is a typical approach whereby wrongly parallelized code is detected and rolled back at runtime. We proposed [27] utilizing processes to implement speculative parallelization using on-stack replacement, allowing for generally simple and portable design where forking a new process enters the speculative state, and killing a faulty process simply performs the roll back operation. While the cost of such operations are high, the approach is promising for cases where the parallel section is long and dependency issues are rare. Also, our proposed system performs speculative parallelization on binary code at runtime, without the need for source code, restarting the program or special hardware support. Initial experiments show about 2× to 3× speedup for speculative execution over serial, when three fourth of loop iterations are parallelizable. Maximum speculation overhead over pure parallel execution is measured at 5.8 %.

This research was partially done within the context of the project PHC IMHOTEP.

7.1.3. Autotuning

Participants: Loïc Besnard, Imane Lasri, Pierre Le Meur, Erven Rohou.

The ANTAREX project relies on a Domain Specific Language LARA [0] of the Clava environment [0]. This DSL is based on Aspect Oriented Programming concepts to allow applications to enforce extra functional properties such as energy-efficiency and performance and to optimize Quality of Service in an adaptive way. The DSL approach allows the definition of energy-efficiency, performance, and adaptivity strategies as well as their enforcement at runtime through application autotuning and resource and power management [28], [29].

In this context, this year we have integrated in Clava some technologies: the memoization, the precision tuning and the loop splitting compilation.

7.1.3.1. Memoization

The concept of memoization essentially involves saving the results of functions together with their inputs so that when the input repeats, the result is taken from a look-up table. This technique, whose objective is to improve sequential performance, has been implemented for C and C++ languages. The support library of this technology allows in particular flexibility for the table management. This work has been submitted for publication in the Elsevier journal SoftwareX. The support library is available at https://gforge.inria.fr/projects/memoization (registered with APP under number IDDN.FR.001.250029.000.S.P.2018.000.10800)

7.1.3.2. Precision tuning

The developed aspects on the type precision consist in the parametrization of the applications in terms of types. Indeed, error-tolerating applications are increasingly common in the emerging field of real-time HPC. Thus, recent works investigated the use of customized precision in HPC as a way to provide a breakthrough in power and performance. This parametrization allows to test easily and quickly different type representations (such as `double`, `float`, `fixed-point`).

7.1.3.3. Loop splitting

The loop splitting technique takes advantage of long running loops to explore the impact of several optimization sequences at once, thus reducing the number of necessary runs. We rely on a variant of loop peeling which splits a loop into into several loops, with the same body, but a subset of the iteration space. New loops execute consecutive chunks of the original loop. We then apply different optimization sequences on each loop independently. Timers around each chunk observe the performance of each fragment. This technique may be generalized to combine compiler options and different implementations of a function called in a loop. It is useful when, for example, the profiling of the application shows that a function is critical in term of time of execution. In this case, the user must try to find the best implementation of their algorithm.

This research is done within the context of the ANTAREX FET HPC collaborative project. The software is being registered with APP.

7.1.4. Hardware/Software JIT Compiler

Participant: Erven Rohou.

In order to provide dynamic adaptation of the performance/energy trade-off, systems today rely on heterogeneous multi-core architectures (different micro-architectures on a chip). These systems are limited to single-ISA approaches to enable transparent migration between the different cores. To offer more trade-offs, we can integrate statically scheduled micro-architecture and use Dynamic Binary Translation (DBT) for task migration. However, in a system where performance and energy consumption are a prime concern, the translation overhead has to be kept as low as possible. We propose Hybrid-DBT [26], an open-source, hardware accelerated DBT system targeting VLIW cores. Three different hardware accelerators have been designed to speed-up critical steps of the translation process. Experimental study shows that the accelerated steps are two orders of magnitude faster than their software equivalent. The impact on the total execution time of applications and the quality of generated binaries are also measured.

Our proposed DBT framework targets the RISC-V ISA, for which both OoO and in-order implementations exist. Our experimental results [37] show that our approach can lead to best-case performance and energy efficiency when compared against static VLIW configurations.

[0] https://web.fe.up.pt/~specs/projects/lara/doku.php

[0] http://specs.fe.up.pt/tools/clava

This work is part of the PhD of Simon Rokicki [22], co-advised by Erven Rohou.

7.1.5. Qubit allocation for quantum circuit compilers

Participants: Sylvain Collange, Marcos Siraichi, Victor Careil.

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. In collaboration with Vinícius Fernandes dos Santos, Fernando Pereira and Marcos Yukio Siraichi at UFMG, 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 was presented at the Code Generation and Optimization (CGO) conference [40].

7.2. Processor Architecture

Participants: Sylvain Collange, Niloofar Charmchi, Kleovoulos Kalaitzidis, Pierre Michaud, Daniel Rodrigues Carvalho, André Seznec, Anita Tino.

7.2.1. Value prediction

Participants: Kleovoulos Kalaitzidis, André Seznec.

For the 1st Championship on Value Prediction (CVP1), we have explored the performance limits of value prediction for small value predictors (8KB and 32KB) in the context of a processor assuming a large instruction window (256-entry ROB), a perfect branch predictor, fetching 16 instructions per cycle, an unlimited number of functional units, but a large value misprediction penalty with a complete pipeline flush at commit on a value misprediction

Our proposition EVES, for Enhanced VTAGE Enhanced Stride, combines two predictor components which do not use on the result of the last occurrence of the instruction to compute the prediction. We use an enhanced version of the VTAGE predictor [11], E-VTAGE. Second, we propose an enhanced version of the stride predictor, E-Stride. E-Stride computes the prediction from the last committed occurrence of the instruction and the number of speculative inflight occurrences of the instruction in the pipeline. The prediction flowing out from E-Stride or E-VTAGE is used only when its confidence is high. A major contribution of this study is the algorithm to assign confidence to predictions depending on the expected benefit/loss from the prediction.

The EVES predictor won the three tracks of CVP1 [39].

7.2.2. Compressed caches

Participants: Daniel Rodrigues Carvalho, Niloofar Charmchi, André Seznec.

Recent advances in research on compressed caches make them an attractive design point for effective hardware implementation for last-level caches. For instance, the yet another compressed cache (YACC) layout [14] leverages both spatial and compression factor localities to pack compressed contiguous memory blocks from a 4-block super-block in a single cache block location. YACC requires less than 2 % extra storage over a conventional uncompressed cache. Performance of LLC is also highly dependent on its cache block replacement management. This includes allocation and bypass decision on a miss as well as replacement target selection which is guided by priority insertion policy on allocation and priority promotion policy on a hit. YACC uses the same cache layout as a conventional set-associative uncompressed cache Therefore the LLC cache management policies that were introduced during the past decade can be transposed to YACC. However, YACC features super-block tags instead of block tags. For uncompressed block, these super-block tags can be used to monitor the reuse behavior of blocks from the same super-block. We introduce the First

In Then First Use Bypass (FITFUB) allocation policy for YACC. With FITFUB, a missing uncompressed block that belongs to a super-block that is already partially valid in the cache is not stored in the cache on its first use, but only on its first reuse if any. FITFUB can be associated with any priority insertion/promotion policy. YACC+FITFUB with compression turned off, achieves an average 6.5%/8% additional performance over a conventional LLC, for single-core/multi-core workloads, respectively. When compression is enabled, the performance benefits associated with compression and FITFUB are almost additive reaching 12.7%/17%. This leads us to call this design the Synergistic cache layout for Reuse and Compression (SRC). SRC reaches the performance benefit that would be obtained with a 4× larger cache, but with less than 2 % extra storage [34].

7.2.3. The Omnipredictor

Participant: André Seznec.

Modern superscalar processors heavily rely on out-of-order and speculative execution to achieve high performance. The conditional branch predictor, the indirect branch predictor and the memory dependency predictor are among the key structures that enable efficient speculative out-of-order execution. Therefore, processors implement these three predictors as distinct hardware components. In [35] we propose the omnipredictor that predicts conditional branches, memory dependencies and indirect branches at state-of-the-art accuracies without paying the hardware cost of the memory dependency predictor and the indirect jump predictor. We first show that the TAGE prediction scheme based on global branch history can be used to concurrently predict both branch directions and memory dependencies. Thus, we unify these two predictors within a regular TAGE conditional branch predictor whose prediction is interpreted according to the type of the instruction accessing the predictor. Memory dependency prediction is provided at almost no hardware overhead. We further show that the TAGE conditional predictor can be used to accurately predict indirect branches through using TAGE entries as pointers to Branch Target Buffer entries. Indirect target prediction can be blended into the conditional predictor along with memory dependency prediction, forming the omnipredictor.

7.2.4. Branch prediction

Participant: Pierre Michaud.

The branch predictor is the keystone of modern superscalar micro-architectures. The TAGE predictor, introduced by André Seznec and Pierre Michaud in 2006, is the most storage-efficient conditional branch predictor known today [16]. Although TAGE is very accurate, it does not exploit its input information perfectly, as significant prediction accuracy improvements are obtained by complementing TAGE with a perceptron-based *statistical corrector* using the same input information [18]. The statistical corrector, even small, makes the whole predictor more complex. We proposed an alternative TAGE-like predictor, called BATAGE, making statistical correction superfluous. BATAGE has the same global structure as TAGE but uses a different tagged-entry format and different prediction and update algorithms. The main reason for TAGE needing statistical correction is the *cold-counter* problem, that is, the fact that recently created tagged entries contain little branch history. To solve the cold-counter problem, we replaced the up-down counter in the tagged entry with two counters counting the *taken* and *not-taken* occurrences separately, and we introduced Bayesian confidence estimation based on Laplace's rule of succession. We also introduced a method called *Controlled Allocation Throttling* for adjusting the rate of creation of tagged entries dynamically. The resulting predictor, BATAGE, obviates the need for external statistical correction [25].

7.2.5. 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 published in the Journal of Parallel and Distributed Computing [6].

7.2.6. Toward out-of-order SIMT micro-architecture

Participants: Sylvain Collange, Anita Tino.

Prior work highlights the continued importance of maintaining adequate sequential performance within throughput-oriented cores [49]. 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.

7.3. WCET estimation and optimization

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

7.3.1. WCET estimation for many core processors

Participants: Imen Fassi, Damien Hardy, Viet Anh Nguyen, Isabelle Puaut, Benjamin Rouxel, Stefanos Skalistis.

7.3.1.1. 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 is 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 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. An implementation on the Kalray MPPA machine allowed to identify implementation-related overheads. All results are described in the PhD thesis document of Viet Anh Nguyen [21], defended in February 2018.

This research is part of the PIA Capacités project.

7.3.1.2. Shared resource contentions and WCET estimation

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.

The work along this direction is part of the PhD thesis of Benjamin Rouxel, defended in December 2018. The new results in 2018 concern scheduling/mapping of parallel applications on multi-core systems using ScratchPad Memories (SPMs). We have recently proposed techniques that jointly select SPM contents off-line, in such a way that the cost of SPM loading/unloading is hidden. Communications are fragmented to augment hiding possibilities. Experimental results show the effectiveness of the proposed techniques on streaming applications and synthetic task-graphs. The overlapping of communications with computations allows the length of generated schedules to be reduced by 4 % in average on streaming applications, and by 8 % in average (with maximum of 16 % for both test cases) for synthetic task graphs. We further show on a case study that generated schedules can be implemented with low overhead on a predictable multi-core architecture (Kalray MPPA) [23].

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

Parallel architectures are nowadays not only 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. The Argo toolchain operates on Scilab and XCoS inputs, and targets ScratchPad Memory (SPM)-based multi-cores. Data-layout and loop transformations play a key role in this flow as they improve SPM efficiency and reduce the number of accesses to shared main memory.

In our most recent work [33], we study how these transformations impact WCET estimates of sequential codes. We demonstrate that they can bring significant improvements of WCET estimates (up to 2.7×) provided that the WCET analysis process is guided with automatically generated flow annotations obtained using polyhedral counting techniques.

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

7.3.2. WCET estimation and optimizing compilers

Participants: Imen Fassi, Isabelle Puaut.

Compiler optimizations, although reducing the execution times of programs, raise issues in static WCET estimation techniques and tools. Flow facts, such as loop bounds, may not be automatically found by static WCET analysis tools after aggressive code optimizations. In this work [36], we explore the use of iterative compilation (WCET-directed program optimization to explore the optimization space), with the objective to (i) allow flow facts to be automatically found and (ii) select optimizations that result in the lowest WCET estimates. We also explore to which extent code outlining helps, by allowing the selection of different optimization options for different code snippets of the application.

7.3.3. Partial WCET

Participants: Rabab Bouziane, Erven Rohou.

Computing the worst-case execution time (WCET) of tasks is important for real-time system design. The industry and research communities have developed a wealth of techniques to compute relevant WCET approximations. Traditionally, WCETs are estimated at the granularity of a function (or task). We propose an approach to estimate partial WCET (δ-WCET), i.e., the worst-case execution time between two locations in a function, such as basic blocks or instructions. Our technique [41] is derived from the well-known implicit path enumeration technique. It takes into account both the control flow graph and the architecture (pipeline and cache hierarchy). Some useful applications of such δ-WCETs are motivated in this paper.

This research is part of the ANR Continuum project.

7.4. Security

Participants: Damien Hardy, Byron Hawkins, Nicolas Kiss, Kévin Le Bon, Erven Rohou.

7.4.1. Compiler-based automation of side-channel countermeasures

Masking is a popular protection against side-channel analysis exploiting the power consumption or electromagnetic radiations. Besides the many schemes based on simple Boolean encoding, some alternative schemes such as Orthogonal Direct Sum Masking (ODSM) or Inner Product Masking (IP) aim to provide more security, reduce the entropy or combine masking with fault detection. The practical implementation of those schemes is done manually at assembly or source-code level, some of them even stay purely theoretical. We propose a compiler extension to automatically apply different masking schemes for block cipher algorithms. We introduce a generic approach to describe the schemes and we manage to insert three of them at compile-time on an AES implementation. A practical side-channel analysis as well as fault injections have been performed on an Arm microcontroller to assess the correctness of the code inserted.

The resulting compiler plugin (sigmask) is registered with APP under number IDDN.FR.001.490003.000.S.P. 2018.000.10000

This research was done within the context of the project ANR CHIST-ERA SECODE.

7.4.2. Program protection through dynamic binary rewriting

Programs written in languages such as C and C++ are prone to memory corruptions because of the manual management of the memory from the programmer. Even today, memory corruptions are among the most dangerous vulnerabilites. According to the MITRE ranking, these bugs are considered one of the top three most dangerous software vulnerabilites.

Thanks to our library Padrone, we are able to instrument the execution of a program with a minimal overhead, making it possible to move or add code in the target process during its execution. We showed that we can change the address of a function at runtime, thus presenting a moving target to an attacker, and making attacks more difficult.

Many security policies have been developed to protect programs. One of them, the Control-Flow Integrity (CFI) ensures the control-flow of the program cannot be altered, preventing the execution of malicious code. Unfortunately, implementations of precise CFI impose a consequent overhead in performance, due to the instrumentation of the execution of the program. We work on building a solution that is able to adapt its protection level to the situation. Adapting the protection level allows us to reduce even further the overhead in performance when the protection is not needed.

AOSTE2 Team

7. New Results

7.1. Uniprocessor Mixed-Criticality Real-Time Scheduling

Participants: Liliana Cucu, Robert Davis, Mehdi Mezouak, Yves Sorel.

In the context of the FUI CEOS project 9.1.1.1 , last year we tranformed the PX4 autopilot free software program in a graph of tasks. In this project our main goal is to perform a real-time schedulability analysis on this program in order to prove that the autopilot will meet all its deadlines when it will operate in the multirotor drone the project is intended to built. The tasks will be executed on a Pixhawk electronic board based on an ARM Cortex M4 microprocessor running on the NuttX OS.

We start by determinating the period and measuring the average execution time of each task which is less than the worst case execution time (WCET). Then, using these periods and these measured execution times we perform an online schedulability analysis using a rate monotonic policy (RM) that shown the set of tasks is not schedulable. Consequently, we informed the partners of the CEOS project that the present version of PX4 is not real-time.

Presently, we are transforming the original set of tasks into a set of real-time tasks. To achieve this goal, we associate to every task a periodic high resolution timer corresponding to the period of the task. Each timer generates an interruption when it expires and the task is put in the ready task queue. The scheduler of NuttX will choose in this queue the task to be executed. In order to validate this transformation we operated the multirotor drone in a simulation tool composed of Gazebo for the geometrical environment of the drone and of the Ground Control Station for setting and controlling the drone. We performed two kinds of simulations, a software in the loop simulation (SitL) which simulates the Pixhawk board, the sensors and the actuators, and a hardware in the loop simulation (HitL) which simulates only the sensors and the actuators, whereas the PX4 program runs on the Pixhawk board. We tested the set of real-time tasks in SitL and we are presently testing them in HitL.

Since we can easily change the period of every task, we plan to modify the periods to make the set of real-time tasks schedulable using an online RM schedulability analysis.

In order to manage high criticality real-time tasks we plan to use an offline scheduler whose scheduling table is generated by an offline schedulability analysis tool that is developped in the team. We plan to modify NuttX in order to support such scheduler.

Finally, in order to complete the real-time schedulability analysis of PX4, we estimate the worst case execution time (WCET) of each task. This problem is 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 statistical reasoning in characterizing the timing behavior (WCET, schedulability analyses) of mixed-criticality real-time applications. The isolated execution times of the programs have indicated large variations indicating expected larger variability in real execution scenarios. In order to decrease the pessimism of the statistical bounds, we are adapting our models to move towards multi-variate approaches.

7.2. Multiprocessor Real-Time Scheduling

Participants: Slim Ben Amor, Evariste Ntaryamira, Salah Eddine Saidi, Yves Sorel, Walid Talaboulma.

The last part of the PhD thesis of Salah Eddine Saidi, was dedicated to the parallelization of FMI-based co-simulation under real-time constraints. More precisely we address HiL (Hardware in the Loop) co-simulation where a part of the co-simulation is replaced by its real counterpart which 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 (resp. ouputs) of the real part are sampled periodically, sending (resp. receiving) data to (resp. from) the simulated part. Every 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 data dependent functions that specify the co-simulation (simulated part). Starting from these constraints we have to schedule the FMI-based co-simulation on a multi-core. We propose an ILP-based algorithm as well as a heuristic that allow the execution of the co-simulation on a multi-core processor while ensuring the previously defined real-time constraints are respected [6]. The proposed heuristic is a list scheduling heuristic. It builds the multi-core schedule iteratively. At each iteration, a list of candidate functions is constructed. The heuristic computes the priority for each candidate function on every core and selects the core for the which the priority is maximized. The priority of a function is a dynamic priority as its computation depends on the partial scheduling solution that has already been computed.

All works achieved by Salah Eddine Saidi on the parallelization of FMI-based co-simulation of numerical models were presented in his PhD thesis defense and manuscript [1].

Avionics applications are based on the specification of “data chains”. Every data chain is a sequence of periodic real-time communicating tasks that are processing the data from sensors up to actuators. Such data chain determines an order in which the tasks propagate data but not in which they are executed. Indeed, inter-task communication and scheduling are independent. We focus on the latency computation, considered as the time elapsed from getting the data from an input and processing it to an output of a data chain. We propose a method for the worst-case latency computation of data chains composed of periodic tasks and executed by a partitioned fixed-priority preemptive scheduler upon a multiprocessor platform [5].

The PhD thesis of Slim Ben Amor is dedicated to the study of multiprocessor scheduling of real-time systems in presence of precedence constraints. This year we have proposed new models [10] for dependent real-time task with probabilistic worst-case execution time (WCET) that are scheduled using a partitioned reasoning. We explore existing solutions from [15] as the closest problem to our dependent task scheduling on multiprocessor and we study their extension to probabilistic models. We conclude that the probabilistic extension would be very difficult with heavy computation since the deterministic solution is based on the resolution of complex ILP optimization problem. Then, we decide to build a new solution to the deterministic problem that should be simple to extend to probabilistic problem. The proposed solution [11] consists of calculating the response time of each sub-tasks in a given DAG task taking in consideration preemptions caused by higher priority sub-tasks executed on the same processor. Then, we evaluate the global response time of the whole graph layer by layer, which allows deciding the schedulability of the entire system.

During the third year of Walid Talaboulma PhD thesis, we continued exploring solutions to make the WCET (Worst Case Execution Time) estimation as independent as possible with respect to the memory accesses. WCET analysis done on a unicore processor (in isolation) is not sufficient when we run our tasks on a multicore processors, the problem of Co-runner interference arises due to contention in shared hardware. Our solution is based on the generation of 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 solution does not necessarily require modifications of software layer, or recompilation of task code. We use those profiles to account for co runners interference and add it to WCET value obtained in isolation, and by updating our schedule, we can also insert idle times at correct scheduling events to decrease the interference.

The PhD thesis of Evariste Ntaryamira is dedicated to the study of multiprocessor real-time systems while ensuring the data freshness. This year we have underlined the difficulty of this scheduling problem [13], [8] while proposing a model to include both time and data constraints. We explore existing solutions from [16] as the closest problem to our data-dependent scheduling problem. The case study associated to this thesis is jointly prepared with the members of the RITS Inria team.

7.3. 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, Safran Aircraft Engines, and Kalray.

The key difficulty of real-time scheduling is that timing analysis and resource allocation depend on each other. An exhaustive search for the optimal solution not being possible for complexity reasons, heuristic approaches are used to break this dependency cycle. Two such approaches are typical in real-time systems design. The first approach uses unsafe timing characterizations for the tasks (e.g., measurements) to build the system, and then checks the respect of real-time requirements through a global timing analysis. The second approach uses a formal model of the hardware platform enabling timing characterizations that are safe for all possible resource allocations (worst-case bounds).

So far, the practicality of the second approach had never been established. Automated real-time parallelization flows still relied on simplified hypotheses ignoring much of the timing behavior of concurrent tasks, communication and synchronization code. And even with such unsafe hypotheses, few studies and tools considered the—harmonic—multiperiodic task graphs of real-world control applications, and the problem of statically managing all their computational, memory, synchronization and communication resources.

This year, we presented the first demonstration of the feasibility of the second approach, showing good practical results for classes of real-world applications and multiprocessor execution platforms whose timing predictability allows keeping pessimism under control. This requires something that is missing in previous work: *the tight orchestration of* **all** *implementation phases*: WCET analysis, resource allocation, generation of *glue code* ensuring the sequencing of tasks on cores and the synchronization and memory coherency between the cores, compilation and linking of the resulting C code. This orchestration is conducted on very detailed timing model that considers both the tasks and the generated glue code, and which includes resource access interferences due to multi-core execution. While orchestration is our main contribution, it should not be understood as a mere combination of existing tools and algorithms. The whole point of our approach is to carefully coordinate every analysis, mapping and code generation phase to enable predictable execution and to keep pessimism under control. To this end, we contributed application normalization phase to facilitate timing analysis, an original code generation algorithm designed to provide mapping-independent worst-case execution time bounds, and new real-time scheduling algorithms capable of orchestrating memory allocation and scheduling.

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. First results are presented in the report [9].

7.4. Real-time Platform Modeling

Participants: Fatma Jebali, Dumitru Potop Butucaru.

This work took place in the framework of the ITEA3 ASSUME project, which funds the post-doc of Fatma Jebali.

One key difficulty in embedded systems design is the existence of multiple models of the same hardware system, developed separately, at different abstraction levels, and used in various phases of the design flow. In the design of real-time embedded systems, we can identify, among other:

- Cycle-accurate system models used to perform fine-grain hardware simulation, mostly during HW and driver design phases. These models provide an exact functional and temporal representation of system execution.

- Microarchitectural models used for pipeline simulation during WCET (*Worst-Case Execution Time*) analysis [19], [20], [18]. These models are used to compute safe over-approximations of the duration of a sequential piece of code, i.e., one function running without interruption on a processor core). To provide precise results, these models preserve much of the microarchitectural detail of processor pipelines and memory hierarchy (e.g. cache states, data transfer latencies).

Both simulation models usually have cyclic activation patterns, but establishing semantic consistency between them is challenging for several reasons. First, the activation pattern, which is the logical time base of the simulation, depends on the abstraction level. In cycle-accurate models, simulation cycles correspond to hardware clock ticks, whereas in WCET analysis models they correspond to changes in the program counter of the sequential program. Second, data abstractions are different in the two simulation models. Cycle-accurate simulators are often also *bit-accurate*, *i.e.* provide exactly the same results as the actual hardware. By comparison, pipeline simulators in WCET analysis abstract away most data types and related operators, typically retaining only Booleans, which can be exploited at analysis time. Last, but not least, the simulators are usually pieces of C/C++ code manually written by different teams or obtained through complex translation processes from high-level Architecture Description Languages (ADLs) that may not have a clear semantics. Formally relating such pieces of code is difficult.

This year we proposed a method to ensure the semantic consistency between the two HW models we consider, focusing on time abstraction issues. Our method relies on *desynchronization* theory [25], which defines sufficient properties ensuring that a synchronous model can be seen as an asynchronous Kahn Process Network (KPN). When a synchronous HW model satisfies these properties, any scheduling of its computations that is compatible with data dependencies will produce the same result (a property known as scheduling-independence). We showed how to control scheduling through changes of the logical time base of the model prior to code generation using a synchronous language compiler. In particular, a careful choice of the logical time base allows us to produce, from the same model, either a cycle-accurate simulator, or the one needed for WCET analysis. In conjunction with some data abstraction, this logical time manipulation allows the synthesis of semantically consistent simulators from a single model.

Furthermore, we can ensure by construction that synchronous models satisfy the properties required by desynchronization theory. To this end, we introduced a new hardware modelling language, named xMAStime, allowing the compositional modeling of systems satisfying the required properties. Results were presented at the ACSD'18 conference [4].

HYCOMES Project-Team

6. New Results

6.1. Hybrid Systems Modeling and Verification

6.1.1. Building a Hybrid Systems Modeler on Synchronous Languages Principles

Participants: Albert Benveniste, Benoît Caillaud.

Hybrid systems modeling languages that mix discrete and continuous time signals and systems are widely used to develop Cyber-Physical systems where control software interacts with physical devices. Compilers play a central role, statically checking source models, generating intermediate representations for testing and verification, and producing sequential code for simulation and execution on target platforms. In [5], Albert Benveniste, Timothy Bourke (PARKAS team Inria/ENS Paris), Benoît Caillaud, Jean-Louis Colaço, Cédric Pasteur (ANSYS/Esterel Technologies, Toulouse) and Marc Pouzet (PARKAS team Inria/ENS Paris) propose a comprehensive study of hybrid systems modeling languages (formal semantics, causality analysis, compiler design, ...). This paper advocates a novel approach to the design and implementation of these languages, built on synchronous language principles and their proven compilation techniques. The result is a hybrid systems modeling language in which synchronous programming constructs can be mixed with Ordinary Differential Equations (ODEs) and zero-crossing events, and a runtime that delegates their approximation to an off-the-shelf numerical solver. We propose an ideal semantics based on non standard analysis, which defines the execution of a hybrid model as an infinite sequence of infinitesimally small time steps. It is used to specify and prove correct three essential compilation steps: (1) a type system that guarantees that a continuous-time signal is never used where a discrete-time one is expected and conversely; (2) a type system that ensures the absence of combinatorial loops; (3) the generation of statically scheduled code for efficient execution. Our approach has been evaluated in two implementations: the academic language Zélus, which extends a language reminiscent of Lustre with ODEs and zero-crossing events, and the industrial prototype Scade Hybrid, a conservative extension of Scade 6.

6.1.2. Structural Analysis of Differential-Algebraic Equations (DAE), State-of-the-Art

Participants: Khalil Ghorbal, Mathias Malandain.

In a deliverable [0] for the FUI ModeliScale collaborative project, Mathias Malandain and Khalil Ghorbal discuss the state-of-the-art methods for performing what is called structural index reduction for differential-algebraic equations, that is equations involving both differential and algebraic equality constraints. Index reduction is one of the basic required methods implemented in any DAE-based modelling language (like Modelica). It is a mandatory step to perform prior to calling a numerical solver to effectively advance time by integrating the set of equations. We cover in particular a recent work that tackles extended models involving several modes, each of which is encoded as a standard DAE.

6.1.3. Multi-Mode DAE Models: Challenges, Theory and Implementation

Participants: Albert Benveniste, Benoît Caillaud, Khalil Ghorbal.

[0] Modeliscale project, deliverable M2.1.1 1, Structural Analysis of Differential-Algebraic Equations (DAE), State-of-the-Art.

The modeling and simulation of Cyber-Physical Systems (CPS) such as robots, vehicles, and power plants often require models with a time varying structure, due to failure situations or due to changes in physical conditions. These are called multi-mode models. In [17], Albert Benveniste, Benoît Caillaud, Hilding Elmqvist (Mogram AB, Lund, Sweden), Khalil Ghorbal, Martin Otter (DLR-SR, Oberpfaffenhofen, Germany) and Marc Pouzet (PARKAS team, Inria/ENS Paris) are interested in multi-domain, component-oriented modeling as performed, for example, with the modeling language Modelica that leads naturally to Differential Algebraic Equations (DAEs). This paper is thus about multi-mode DAE systems. In particular, new methods are introduced to overcome one key problem that was only solved for specific subclasses of systems before: How to switch from one mode to another one when the number of equations may change and variables may exhibit impulsive behavior? An evaluation is performed both with the experimental modeling and simulation system Modia, a domain specific language extension of the programming language Julia, and with SunDAE, a novel structural analysis library for multi-mode DAE systems.

6.1.4. Vector Barrier Certificates and Comparison Systems

Participant: Khalil Ghorbal.

Vector Lyapunov functions are a multi-dimensional extension of the more familiar (scalar) Lyapunov functions, commonly used to prove stability properties in systems of non-linear ordinary differential equations (ODEs). In [7], Kahlil Ghorbal and Andrew Sogokon (CMU, Pittsburgh, USA) explore an analogous vector extension for so-called barrier certificates used in safety verification. As with vector Lyapunov functions, the approach hinges on constructing appropriate comparison systems, i.e., related differential equation systems from which properties of the original system may be inferred. The paper presents an accessible development of the approach, demonstrates that most previous notions of barrier certificate are special cases of comparison systems, and discusses the potential applications of vector barrier certificates in safety verification and invariant synthesis.

6.2. Contract-based Reasoning for Cyper-Physical Systems Design

6.2.1. Contracts for Cyper-Physical Systems Design

Participants: Albert Benveniste, Benoît Caillaud.

Contract-based reasoningn has been proposed as an “orthogonal” approach that complements methodologies proposed so far to cope with the complexity of cyber-physical systems design. Contract-based reasoning provides a rigorous framework for the verification, analysis, abstraction/refinement, and even synthesis of cyber-physical systems. A number of results have been obtained in this domain but a unified treatment of the topic that can help put contract-based design in perspective was missing. In [6], Albert Benveniste, Benoît Caillaud and co-authors provide a unified theory where contracts are precisely defined and characterized so that they can be used in design methodologies with no ambiguity. This monograph gathers research results of the former S4 inria team. It identifies the essence of complex system design using contracts through a *mathematical meta-theory*, where all the properties of the methodology are derived from an abstract and generic notion of contract. We show that the meta-theory provides deep and enlightening links with existing contract and interface theories, as well as guidelines for designing new theories. Our study encompasses contracts for both software and systems, with emphasis on the latter. We illustrate the use of contracts with two examples: requirement engineering for a parking garage management, and the development of contracts for timing and scheduling in the context of the Autosar methodology in use in the automotive sector.

6.2.2. Cyber-Physical Systems Design: from Natural Language Requirements

In his current PhD work, co-supervised by Benoît Caillaud and Annie Forêt (SemLIS, IRISA, Rennes, France), Aurélien Lamercerie explores the construction of formal representations of natural language texts. The mapping from a natural language to a logical representation is realized with a grammatical formalism, linking the syntactic analysis of the text to a semantic representation. In [44], Aurélien Lamercerie targets behavioral specifications of cyber-physical systems, ie any type of system in which software components interact closely with a physical environment. The objective is the simulation and formal verification, by automatic or assisted methods, of system level requirements expressed in a controled fragment of a natural language.

KAIROS Team

7. New Results

7.1. Schedulability of CCSL specifications via SMT

Participants: Frédéric Mallet, Robert de Simone.

The full expressive power of the CCSL language makes it very complex, if not impossible, to also find good, or even optimal, schedules as results of solving the CCSL constraints. Nevertheless, important subclasses can be devised, or efficient heuristics can be attempted. The study of CCSL scheduling decidability and efficient is a long-term source of theoretical developments in the team, here is a record of this year advances, split in two parts.

We have made progress on the inherent complexity of finding a schedule with a general CCSL specification. We have proved that the schedulability problem of CCSL is NP-hard. Then it makes sense to find whether there are still some practical ways to find solutions in specific cases. It turns out that in many cases, we can still find solutions in a reasonable duration. To do so, we have proposed [8] an encoding of CCSL specifications as an SMT (Satisfiability Modulo Theory) specification and we use Z3 and CVC4 as solvers for our experiments. Using a pure SAT solver is not possible for CCSL, as CCSL combines Boolean operations with arithmetics on unbounded integers. Using SMT allows to combine both. This encoding uses a sublogic called UFLIA that relies on quantified variables (boolean or integer), undefined functions on boolean and integers, and linear integer arithmetics. This logics is undecidable in the general case and the use of quantified variables makes it difficult to deal with, but we have found some interesting examples where we still get some results in a reasonable amount of time. We have also tried to identify subdomains where we get interesting results and we have focused on pure real-time schedulability problems. In that context, we showed that the schedulability problem for a set of real-time tasks reduces to the schedulability problem of CCSL specifications with a specific form (to be published).

The Clock Constraint Specification Language (CCSL) is a clock-based specification language for capturing causal and chronometric constraints between events in Real-Time Embedded Systems (RTESs). Due to the limitations of the existing verification approaches, CCSL lacks a full verification support for 'unsafe CCSL specifications' and a unified proof framework. In this paper [18], we propose a novel verification approach based on theorem proving and SMT-checking. We firstly build a logic called CCSL Dynamic Logic (CDL), which extends the traditional dynamic logic with 'signals' and 'clock relations' as primitives, and with synchronous execution mechanism for modelling RTESs. Then we propose a sound and relatively complete proof system for CDL to provide the verification support. We show how CDL can be used to capture RTES and verify CCSL specifications by analyzing a simple case study.

7.2. Logical Time for the semantics of Reactive Languages

Participants: Frédéric Mallet, Robert de Simone.

This work was initiated during the sabbatical period of Reihard von Hanxleden, on leave from the University of Kiel (Germany), funded by the the UMR I3S laboratory.

The results won Best Paper Award at the Federated Design Languages (FDL) conference edition of 2018 [16]. The paper abstract follows:

Synchronous languages, such as the recently proposed SCCharts language, have been designed for the rigorous specification of real-time systems. Their sound semantics, which builds on an abstraction from physical execution time, make these languages appealing, in particular for safety-critical systems. However, they traditionally lack built-in support for physical time. This makes it rather cumbersome to express things like time-outs or periodic executions within the language. We here propose several mechanisms to reconcile the synchronous paradigm with physical time. Specifically, we propose extensions to the SCCharts language to express clocks and execution periods within the model. We draw on several sources, in particular timed automata, the Clock Constraint Specification Language, and the recently proposed concept of dynamic ticks. We illustrate how these extensions can be mapped to the SCChart language core, with minimal requirements on the run-time system, and we argue that the same concepts could be applied to other synchronous languages such as Esterel, Lustre or SCADE.

7.3. Dealing with uncertainty in logical time

Participants: Frédéric Mallet, Robert de Simone.

When uplifting the target of models to heterogeneous Cyber-Physical Systems, the relations from physical time (which governs Physical components) to logical time becomes an issue for proper abstraction in the design. Often, the other engineering discipline may know of "proto-logical" timing abstraction, but involving probabilistic/stochastic ingredients to link the declared logical clocks/events. As a results, several attempts have been made at extending the language to allow perceptive probabilistic structuring operators, that may link (unreachable) physical rhythms with their discretized, manageable counter-parts. Of course the feasibility of constraint solving remains the key issue for allowing extensions scarcely. Nevertheless, it should be noted that the focus on relevancy of relations between physical and logical times may in some case be an important concerns for non-IT scientists.

The reports on how early attempts can be found in [6], [10], [12]. The topic is far from closed, but as such these are valuable starts.

In the future, we plan to exploit these model extensions on practical application fields, including car trajectory computation with Renault Software Lab, security properties "with Time"in the ILP SPAI with other Inria teams, and micro-satellites in the ATIPPIC IRT Saint-Exupery project with Thales Alenia Space.

7.4. Behavioral semantics and equivalence notions for Open Systems

Participants: Eric Madelaine, Tengfei Li, Zechen Hou.

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. 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. Following last year results we have published an experience paper [23] showing the applicability of this approach to show properties of a piece of the control software of a nano-satellite, specified using BIP architectures. Our work now turns on designing specific symbolic algorithms for model checking and equivalence checking (bisimulation) of such open systems, and also, as a specific application domain, to formalize the encoding of BIP architecture, extended with data constraints, into open pNets, aiming at a full approach for compositional verification of such systems. This work is done in collaboration with researchers from ENS Lyon and Inria Lille, and from ECNU Shanghai [23].

7.5. Logical Time for Safety Analysis and dependability

Participants: Paul Bouche, Amin Oueslati, Robert de Simone.

We have studied in the past the relevance of Logical Time for modeling of dynamic Non-Functional Properties (NFP) aspects of functional applications and/or execution platforms. In this setting, any recurring events may be seen as generating its own "rythm", as a logical clock. The most obvious NFP aspects to consider were performance and power consumption, as important concerns of Real-Time Embedded systems. Recently we have turned towards fault tolerance and availability/dependability aspects. This was motivated by demands from industrial partners inside IRT Saint-Exupery, who tried to design in real terms the digital computing structure of micro-satellites using ordinary processor components from the Shelf (COTS), extremely sensible to solar radiations (creatings faults). We have put up a full model-based design of the proposed use case, which includes modeling of the fault-tolerant features, but also the independent modeling of waterfall propagation schemes from incidental faults to fully recognized dysfunctions, where the system is no longer operational. Current results are encouraging, as they build up natural specification styles using logical time on top of existing formalisms such as AltaRica, widely used in industry. Methodological advances are proposed to industrial partners in IRT Saint-Exupery, and primarily Thales Alenia Space. We plan to comfort our approach next year with dedicated tools for modeling and analysis, as well as translation towards existing formalisms such as AltaRica, seen as lower level in our context.

7.6. Co-Simulation of Cyber-Physical Systems

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

While we continued to study and envision the past, present and future of co-simulation in [11], we already obtained promising results. In [14], we highlighted the current problems of the FMI co-simulation standards and more generally of existing coordination between actors of the co-simulation. We also shown that providing appropriate mean to communicate with the actors according to their internal semantics allows for dedicated coordinator providing better results than existing ones (speed up can reach 25 with a perfect accuracy). As shown in [14], the functional correctness of co-simulation can be violated by a non appropriate coordination of co-simulation actors. To avoid such phenomenon, we explored in [17] the possibility to formally prove the correctness of a coordinator according to properties defined by the actors. This last work is greatly exploratory but Julien Deantoni did a Short Term Scientific Mission (in the context of the MPM4CPS cost action [0]) in the MSDL Lab in Antwerp to understand more deeply the problem and potential solutions. Preliminary interesting results have been obtained [0] and may be published in 2019.

7.7. Early Interconnect Contention Analysis

Participants: Amin Oueslati, Julien Deantoni.

In the context of the Atippic project, industrial partners are using the Capella system engineering language (http://polarsys.org/capella) to migrate a satellite control software on a totally new architecture platform based on "COTS" dual core processors. In order to better deal with the potential contention on the interconnect between the different cores, it was required to help for contention analysis. In this context and based on one of our software (GEMOC Studio: http://eclipse.org/gemoc) we developed an executable extension to Capella, from which simulation of Capella model can be used to obtain bus latency and bandwidth.

We are currently extending this simulation approach to ease Design Space Exploration based on variation of some parameters (typically parameters of the tasks that create traffic like for instance, periods or consumed/produced data size). First results have already been demonstrated to the IRT Saint-Exupery and should be published early 2019.

7.8. Process network models with explicit data size handling

Participants: Amin Oueslati, Robert de Simone.

[0] http://mpm4cps.eu/

[0] http://mpm4cps.eu/STSM/reports/material/STSM_DeantoniJulien_Report_527.pdf

We concluded our activities in the definition of a process network, inspired from established formalisms such as Ptolemy's SDF, StreaMIT, and Thales Array-OL task graph languages. Our next formalisms described accurately how regular data structures (2-dimensional arrays or matrices mostly) get assembled or deassembled in actual data-flow computations for streaming intensive data/signal processing. This allows to allocate these computations to similar dedicated architectures (GPUs, TPUs) while making all kinds of parallelism (data-, task-, streaming) explicit. The resulting forms of specification are intently very close to representations that may be expressed in OpenMP or MPI, and cover the important class of Deep Networks filter stream models, which have raised tremendous interest lately in Artificial Intelligence.

7.9. Union and Intersection constraints

Participants: Luigi Liquori, Claude Stolze.

In [21], we introduced an explicitly typed λ-calculus with strong pairs, projections and explicit type coercions. The calculus can be parameterized with different intersection type theories, producing a family of calculi with related intersection typed systems. We proved the main properties like Church-Rosser, unicity of type, subject reduction, strong normalization, decidability of type checking and type reconstruction. We stated the relationship between the intersection type assignment systems and the corresponding intersection typed systems by means of an essence function translating an explicitly typed Delta-term into a pure λ-term one. We finally translated a term with type coercions into an equivalent one without them; the translation is proved to be coherent because its essence is the identity. The resulting generic calculus can be parametrized to take into account other intersection type theories as the ones in the Barendregt *et al.* book.

7.10. Logical frameworks with Union and Intersection constraints and Oracles

Participants: Luigi Liquori, Claude Stolze.

In [13], we introduced the Δ-framework, DLF, a dependent type theory based on the Edinburgh Logical Framework LF, extended with the *strong proof-functional connectives*, i.e. strong intersection, minimal relevant implication and strong union. Strong proof-functional connectives take into account the shape of logical proofs, thus reflecting polymorphic features of proofs in formulæ. This is in contrast to classical or intuitionistic connectives where the meaning of a compound formula depends only on the truth value or the provability of its subformulæ. Our framework encompasses a wide range of type disciplines. Moreover, since relevant implication permits to express subtyping, DLF subsumes also Pfenning's refinement types. We discuss the design decisions which have led us to the formulation of DLF, study its metatheory, and provide various examples of applications. Our strong proof-functional type theory can be plugged in existing common interactive proof assistants.

Moreover, in [7], we introduced two further extensions of LF, featuring monadic *locks*. A lock is a monadic type construct that captures the effect of an *external call to an oracle*. The oracle can be invoked either to check that a constraint holds or to provide a suitable witness. Such calls are the basic tool for *plugging-in*, i.e. gluing together, different type theories and proof development environments.

7.11. Object reclassification

Participant: Luigi Liquori.

In [19], we investigated, in the context of *functional prototype-based languages*, a calculus of objects which might extend themselves upon receiving a message, a capability referred to by Cardelli as a *self-inflicted* operation. We introduced a sound type system for this calculus which guarantees that evaluating a well-typed expression will never yield a *message-not-found* run-time error. The resulting calculus is an attempt towards the definition of a language combining the safety advantage of static type checking with the flexibility normally found in dynamically typed languages.

7.12. Object discovery

Participant: Luigi Liquori.

In [20], we proposed a Content Name System (CNS) discovery service, extending the current TCP/IP hourglass Internet architecture, that provides a new network aware content discovery service. Contents are addressed using “hypernames", whose rich syntax allow to specify hosts, PKI, fingerprint and optional logical attributes (tags) attached to the content name, such as e.g. mutable vs. immutable contents, digital signatures, owner, availability, price, etc. The CNS behavior and architecture is, partly, inspired by the Domain Name Service (DNS), and whose discovery process logic uses the Border Gateway Protocol (BGP) information allowing Internet to route between different Autonomous Systems (AS). The service registers and discovers object names in each Autonomous System (AS), and the content discovery process is inspired to the so called “valley-free” property. In the routing among different ASes (i.e., the BGP protocol) this is a property that avoids unjustified AS transit costs.

7.13. Code optimization for HPC and CPS programs

Participants: Sid Touati, Carsten Bruns, Robert de Simone.

Optimising HPC applications is a classical research area in computer science, complementary to intensive computation (which is an adjacent research community to HPC). Since decades, the most used languages are imperative ones (FORTRAN, C, etc). These languages are the closest to formal algorithms and low-level assembly codes. In intensive computing area, other kinds of languages and programming paradigms are used (interpreted languages for instance), but are far from HPC challenges, which tackle low level optimization (close to back-end compilation and processor micro-architectures).

We started a while ago to work on optimisation of HPC applications at C++ program level, where code and data are mixed in the same objects, allowing sophisticated programming methods that were not traditionally tackled in classical HPC programming (such as virtual classes, exception handling, etc). Currently, we are working on performance analysis and optimisation of linear algebra codes (BLAS) programmed with classes: this allows to extend BLAS computation to any kind of data (such as complex numbers) not only floating points. Our final aim is to apply and adjust this type of general C++ code optimization, to cover the spectrum of typical Kairos applications expressed from in C++ from high level formal specifications.

PARKAS Project-Team

6. New Results

6.1. 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 modeling 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 new semantic model for the modular reset feature provided by the Scade language and required for the compilation of hierarchical state machines. This work was presented at the SCOPES workshop in Germany in May [17]. Work continues on connecting this semantic model to the intermediate compilation target.

We completed work on generalizing the compiler to treat clocked arguments. This involved changes to our intermediate Obc language and the addition of a pass to add some (necessary) variable initializations in an efficient way. This work was accepted for presentation at the Journées Francophones des Langages Applicatifs in 2019.

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. This has been published in [15]. We are pursuing this line of research studying of the algorithm advances of Julia can be integrated into other programming languages.

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. This has been published in [18].

6.4. Fast and reliable unwinding via DWARF tables

Participants: Theophile Bastian, Francesco Zappa Nardelli.

DWARF is a widely-used debugging data format. DWARF is obviously relied upon by debuggers, but it plays an unexpected role in the runtime of high-level programming languages and in the implementation of program analysis tools. The debug information itself can be pervaded by subtle bugs, making the whole infrastructure unreliable. In this project we are investigating techniques and tools to perform validation and synthesis of the DWARF stack unwinding tables, to speedup DWARF-based unwinding, as well as exploring adventurous projects that can be built on top of reliable DWARF information.

At the time of writing, we have a tool that can validate DWARF unwind tables generated by mainstream compilers; the approach is effective, we found a problem in Clang table generation and several in GLIBC inline-assembly snippets. We also designed and implemented a tool that can synthesise DWARF unwind tables from binary that lacks them (e.g. because the compiler did not generate them - immediate applications: JITs assembly, inline assembly, ...). Additionally we have designed and implemented a ahead-of-time compiler of DWARF unwind tables to assembly, and an ad-hoc unwinder integrated with the defacto standard unwinder libuwind. It can speed up unwinding by a factor between 25x and 60x (depending on application), with a 2.5x size overhead for unwind information.

Discussion is in progress to get these tools included in mainstream tool (e.g. the GNU profiler Perf).

6.5. Sundials/ML: OCaml interface to Sundials Numeric Solvers

Participants: Timothy Bourke, Marc Pouzet.

This year we made major updates to the Sundials/ML OCaml interface to support v3.1.x of the Sundials Suite of numerical solvers.

This release adds support for the new generic matrix and linear solver interfaces. Major work was required to add these new modules, update the existing solver interfaces, and ensure backwards compatibility with Sundials to v2.7.0 (which is still the version installed by Debian stable). We also improved our treatment of integer types used in indexing, refactored the Dls and Sls matrix modules, improved our generation of performance stats (by adding confidence intervals), made the configure script more robust, and untangle the mass-solver and Jacobian interfaces of the ARKODE solver.

6.6. Zélus

Participants: Timothy Bourke, Marc Pouzet.

This year, we made a major revision of the language and compiler, called now the version 2. The language now deal with higher order functions. All the static analyses, type inference, causality inference and the initialization analysis has been extended. The code generation has also been improved, in particular the interface with the numeric solver. Several larger examples have been written.

A paper that present the overall approach followed in ZELUS has been published [12].

6.7. Deterministic Concurrency: A Clock-Synchronised Shared Memory Approach

Participant: Marc Pouzet.

Synchronous programming (SP) provides deterministic concurrency. So far, however, communication has been constrained to a set of primitive clock-synchronised shared memory (scm) data types, such as data-flow registers, streams and signals with restricted read and write accesses that limit modularity and behavioural abstractions. In the paper [23], we propose an extension to the SP theory which retains the advantages of deterministic concurrency, but allows communication to occur at higher levels of abstraction than currently supported by SP data types. Our approach is as follows. To avoid data races, each csm type publishes a policy interface for specifying the admissibility and precedence of its access methods. Each instance of the csm type has to be policy-coherent, meaning it must behave deterministically under its own policy—a natural requirement if the goal is to build deterministic sys- tems that use these types. In a policy-constructive system, all access methods can be scheduled in a policy-conformant way for all the types without deadlocking. In this paper, we show that a policy-constructive program exhibits deterministic concurrency in the sense that all policy-conformant interleavings produce the same input-output behaviour. Policies are conservative and support the csm types existing in current SP languages. This work is a follower of a old work we did in 2009, published at LCTES about scheduling policies.

6.8. Compiling synchronous languages for multi-processor implementations

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

This work was performed with industrial partners in the context of the ASSUME project.

We have continued to improve our front-end tools for a use case provided by Airbus. This tool now generates three kinds of monolithic Lustre program, which are taken as an input of the Lopht tool (AOSTE team), which in turn generates an executable for the Kalray MPPA. In particular, one of the code generators is based on the hyper-period-expansion transformation, which unrolls the computation and generates a single step function running at the slowest period. This transformation allows managing the multi-periodic aspect of the application at the source level. Together with the work of the AOSTE team and Airbus, it allows us to execute the full application on a MPPA (TRL-5 Airbus certification level).

We have also improved the front-end tools for the use case provided by Safran. These tools ware integrated into the Heptagon compiler. Many improvements to the parser and many convenient program transformations (tuple and array destruction, equation clustering, ...) were implemented in the Heptagon compiler in order to treat this use case and enable the Lopht tool to extract the best performance. In particular, we have investigated the impact of inlining on the degree of parallelism exposed by the use-case application.

In addition to the work described above, we have defined a language extension for 1-synchronous clocks, strictly periodic clocks with a single activation. We show that we can derive a scheduling problem from the clock constraints in a program. However, solving these constraints by using an interesting cost functions (such as WCET load balancing across the different phases of a period) with an ILP does not scale for the two use cases. Thus, we used the fact that we do not need the optimal solution to fall back on heuristics, which finds a good solution within acceptable bounds. We have also investigated the effect of a non-determinism operator on the scheduling constraints, which gives extra freedom for choosing a schedule.

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

In collaboration with ANSYS, a compilation technique has been designed for compiling SCADE to multicore [24].

SPADES Project-Team

6. New Results

6.1. Design and Programming Models

Participants: Pascal Fradet, Alain Girault, Gregor Goessler, Xavier Nicollin, Christophe Prévot, Sophie Quinton, Arash Shafiei, Jean-Bernard Stefani, Martin Vassor, Souha Ben Rayana.

6.1.1. A multiview contract theory for cyber-physical system design and verification

The design and verification of critical cyber-physical systems is based on a number of models (and corresponding analysis techniques and tools) representing different viewpoints such as function, timing, security and many more. Overall correctness is guaranteed by mostly informal, and therefore basic, arguments about the relationship between these viewpoint-specific models. More precisely, the assumptions that a viewpoint-specific analysis makes on the other viewpoints remain mostly implicit, and whenever explicit they are handled mostly manually. In [11], we argue that the current design process over-constrains the set of possible system designs and that there is a need for methods and tools to formally relate viewpoint-specific models and corresponding analysis results. We believe that a more flexible contract-based approach could lead to easier integration, to relaxed assumptions, and consequently to more cost efficient systems while preserving the current modelling approach and its tools.

The framework we have in mind would provide viewpoint specific contract patterns guaranteeing inter-viewpoint consistency in a flexible manner. At this point, most of the work remains to be done. On the application side, we need a more complete picture of existing inter-viewpoint models. We also need the theory required for the correctness proofs, but it should be based on the needs on the application side.

6.1.2. End-to-end worst-case latencies of task chains for flexibility analysis

In collaboration with Thales, we address the issue of change during design and after deployment in safety-critical embedded system applications. More precisely, we focus on timing aspects with the objective to anticipate, at design time, future software evolutions and identify potential schedulability bottlenecks. The work presented in this section is the PhD topic of Christophe Prévot, in the context of a collaboration with Thales TRT, and our algorithms are being implemented in the Thales tool chain, in order to be used in industry.

This year, we have completed our work on the analysis of end-to-end worst-case latencies of task chains [10] that was needed to extend our approach for quantifying the flexibility, with respect to timing, of real-time systems made of chains of tasks. In a nutshell, flexibility is the property of a given system to accommodate changes in the future, for instance the modification of some of the parameters of the system, or the addition of a new task in the case of a real-time system.

One major issue that hinders the use of performance analysis in industrial design processes is the pessimism inherent to any analysis technique that applies to realistic system models (*e.g.*, , systems with task chains). Indeed, such analyses may conservatively declare unschedulable systems that will in fact never miss any deadlines. The two main avenues for improving this are (i) computing tighter upper bounds on the worst-case latencies, and (ii) measuring the pessimism, which requires to compute also guaranteed lower bounds. A lower bound is guaranteed by providing an actual system execution exhibiting a behavior as close to the worst case as possible. As a first step, we focus in [10] on uniprocessor systems executing a set of sporadic or periodic hard real-time task chains. Each task has its own priority, and the chains are scheduled according to the fixed-priority preemptive scheduling policy. Computing the worst-case end-to-end latency of each chain is complex because of the intricate relationship between the task priorities. Compared to state of the art analyses, we propose here tighter upper bounds, as well as lower bounds on these worst-case latencies. Our experiments show the relevance of lower bounds on the worst-case behavior for the industrial design of real-time embedded systems.

Based on our end-to-end latency analysis for task chains, we have also proposed an extension of the concept of slack to task chains and shown how it can be used to perform flexibility analysis and sensitivity analysis. This solution is particularly relevant for industry as it provides means by which the system designer can anticipate the impact on timing of software evolutions, at design time as well as after deployment.

6.1.3. Location graphs

We have introduced the location graph model [58] as an expressive framework for the definition of component-based models able to deal with dynamic software configurations with sharing and encapsulation constraints. We have completed a first study of the location graph behavioral theory (under submission), initiated its formalization in Coq, and an implementation of the location framework with an emphasis of the expression of different isolation and encapsulation constraints.

We are now studying conservative extensions to the location graph framework to support the compositional design of heterogeneous hybrid dynamical systems and their attendant notions of approximate simulations [60].

In collaboration with the Spirals team at Inria Lille – Nord Europe, we have applied the location framework for the definition of a pivot model for the description of software configurations in a cloud computing environment. We have shown how to interpret in our pivot model several configuration management models and languages including TOSCA, OCCI, Docker Compose, Aeolus, OpenStack HOT.

6.1.4. 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 (MoCs) [40], [31], we have written a survey providing a comprehensive description of the existing parametric dataflow MoCs [34], and we have studied *symbolic* analyses of dataflow graphs [35]. More recently, we have proposed an original method to deal with lossy communication channels in dataflow graphs [39].

We are now studying models allowing dynamic reconfigurations of the *topology* of the dataflow graphs. In particular, many modern streaming applications have a strong need for reconfigurability, for instance to accommodate changes in the input data, the control objectives, or the environment.

We have proposed a new MoC called Reconfigurable Dataflow (RDF) [15]. RDF extends SDF with transformation rules that specify how the topology and actors of the graph may be reconfigured. Starting from an initial RDF graph and a set of transformation rules, an arbitrary number of new RDF graphs can be generated at runtime. The major quality of RDF is that it can be statically analyzed to guarantee that all possible graphs generated at runtime will be connected, consistent, and live. This is the research topic of Arash Shafiei's PhD, in collaboration with Orange Labs.

6.1.5. Monotonic prefix consistency in distributed systems

We have studied the issue of data consistency in distributed systems. Specifically, we have considered a distributed system that replicates its data at multiple sites, which is prone to partitions, and which is assumed to be available (in the sense that queries are always eventually answered). In such a setting, strong consistency, where all replicas of the system apply synchronously every operation, is not possible to implement. However, many weaker consistency criteria that allow a greater number of behaviors than strong consistency, are implementable in available distributed systems. We have focused on determining the strongest consistency criterion that can be implemented in a convergent and available distributed system that tolerates partitions, and we have shown that no criterion stronger than Monotonic Prefix Consistency (MPC [61], [44]) can be implemented [18].

6.2. Certified Real-Time Programming

Participants: Pascal Fradet, Alain Girault, Gregor Goessler, Xavier Nicollin, Sophie Quinton, Xiaojie Guo, Maxime Lesourd.

6.2.1. Time predictable programming languages and architectures

Time predictability (PRET) is a topic that emerged in 2007 as a solution to the ever increasing unpredictability of today's embedded processors, which results from features such as multi-level caches or deep pipelines [37]. For many real-time systems, it is mandatory to compute a strict bound on the program's execution time. Yet, in general, computing a tight bound is extremely difficult [64]. The rationale of PRET is to simplify both the programming language and the execution platform to allow more precise execution times to be easily computed [27].

We have extended the PRET-C compiler [25] in order to make it energy aware. To achieve this, we use dynamic voltage and frequency scaling (DFVS) and we insert DVFS control points in the control flow graph of the PRET-C program. Several difficulties arise: (i) the control flow graph is concurrent, (ii) the resulting optimization problem is a time and energy multi-criteria problem, and (iii) since we consider PRET-C programs, we actually address the Worst-Case Execution Time (WCET) and the Worst-Case Energy Consumption (WCEC). Thanks to a novel ILP formulation and to a bicriteria heuristic, we are able to address the two objectives jointly and to compute, for each PRET-C program, the Pareto front of the non-dominated solutions in the 2D space (WCET,WCEC) [63]. We have recently improved this result to reduce the complexity of the algorithm and to produce the *optimal* Pareto front. This is the topic of Jia Jie Wang's postdoc.

Moreover, within the CAPHCA project, we have proposed a new approach for predictable inter-core communication between tasks allocated on different cores. Our approach is based on the execution of synchronous programs written in the FOREC programming language on deterministic architectures called PREcision Timed. The originality resides in the time-triggered model of computation and communication that allows for a very precise control over the thread execution. Synchronisation is done via configurable Time Division Multiple Access (TDMA) arbitrations (either physical or conceptual) where the optimal size and offset of the time slots are computed to reduce the inter-core synchronization costs. Results show that our model guarantees time-predictable inter-core communication, the absence of concurrent accesses (without relying on hardware mechanisms), and allows for optimized execution throughput. This is the topic of Nicolas Hili's postdoc.

6.2.2. 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 [4]. A weakly-hard real-time guarantee specifies an upper bound on the maximum number m of deadline misses of a task in a sequence of k consecutive executions. Based on our previous work on Typical Worst-Case Analysis [4], [8], we have introduced in [13] the first verification method which is able to provide weakly-hard real-time guarantees for tasks and task chains in systems with multiple resources under partitioned scheduling with fixed priorities. All existing weakly-hard real-time verification techniques are restricted today to systems with a single resource. Our verification method is applied in the context of switched networks with traffic streams between nodes, and we demonstrate its practical applicability on an automotive case study.

6.2.3. 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.*, [60] 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 [56]. 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 [42]. 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 are currently investigating an approach using mode sequences 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.

6.2.4. 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 [65], *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 [65]. 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 (under a finite horizon), and that the optimal speed at any time t will be a deterministic function of the current state at time t [41]. Under an infinite horizon, we use a *Value Iteration* algorithm.

This work led us to compare several existing speed policies with respect to their feasibility. Indeed, the policies (OA) [65], (AVR) [65], and (BKP) [29] all assume that the maximal speed S_{max} available on the processor is infinite, which is an unrealistic assumption. For these three policies and for our (MDP) policy, we have established necessary and sufficient conditions on S_{max} guaranteeing that no job will ever miss its deadline.

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 real-time systems analyses. Specifically, we contribute to Prosa [23], a Coq library of reusable concepts and proofs for real-time systems analysis. A key scientific challenge is to achieve a modular structure of proofs, *e.g.*, for response time analysis. Our goal is to use this library for:

1. a better understanding of the role played by some assumptions in existing proofs;
2. a formal verification and comparison of different analysis techniques; and
3. the certification of results of existing (*e.g.*, industrial) analysis tools.

Our first major result [16] is a task model that generalizes the digraph model [59] and its corresponding analysis for fixed-priority scheduling with limited preemption. The motivation for this work, which is not yet fully proven in Coq, is to obtain a formally verified schedulability analysis for a very expressive task model. In the context of computer assisted verification, it permits to factorize the correctness proofs of a large number of analyses. The digraph task model seems a good candidate due to its powerful expressivity. Alas, its ability to capture dependencies between arrival and execution times of jobs of different tasks is very limited. Our extended model can capture dependencies between jobs of the same task as well as jobs of different tasks. We provide a correctness proof of the analysis that is written in a way amenable to its formalization in the Coq proof assistant. Despite being much more general, the Response Time Analysis (RTA) for our model is not significantly more complex than the original one. Also, it underlines similarities between existing analyses, in particular the analysis for the digraph model and Tindell's offset model [62].

A second major result is CertiCAN, a tool produced using Coq for the formal certification of CAN analysis results. Result certification is a process that is light-weight and flexible compared to tool certification, which makes it a practical choice for industrial purposes. The analysis underlying CertiCAN is based on a combined use of two well-known CAN analysis techniques [62] that makes it computationally efficient. Experiments demonstrate that CertiCAN is able to certify the results of RTaW-Pegase, an industrial CAN analysis tool, even for large systems. This result paves the way for a broader acceptance of formal tools for the certification of real-time systems analysis results. Beyond CertiCAN, we believe that this work is significant in that it demonstrates the advantage of result certification over tool certification for the RTA of CAN buses. In addition, the underlying technique can be reused for any other system model for which there exist RTAs with different levels of precision. This work will be presented at RTAS 2019.

In parallel, we have completed and published in [17] a Coq formalization of Typical Worst-Case Analysis (TWCA) [4], [8], an analysis technique for weakly-hard real-time systems. Our generic analysis is based on an abstract model that characterizes the exact properties needed to make TWCA applicable to any system model. Our results are formalized and checked using the Coq proof assistant along with the Prosa schedulability analysis library. This work opens up new research directions for TWCA by providing a formal framework for the trade-off that must be found between time efficiency and precision of the analysis. Hopefully, our generic proof will make it easier to extend TWCA to more complex models in the future. In addition, our experience with formalizing real-time systems analyses shows that it is not only a way to increase confidence in the results of the analyses; it also helps understanding their key intermediate steps, the exact assumptions required, and how they can be generalized.

6.2.6. Logical execution time

In collaboration with TU Braunschweig and Daimler, we have worked on the application of the Logical Execution Time (LET) paradigm [50], according to which data are read and written at predefined time instants, to the automotive industry. The LET paradigm was considered until recently by the automotive industry as not efficient enough in terms of buffer space and timing performance. The shift to embedded multicore processors has represented a game changer: The design and verification of multicore systems is a challenging area of research that is still very much in progress. Predictability clearly is a crucial issue which cannot be tackled without changes in the design process. Several OEMs and suppliers have come to the conclusion that LET might be a key enabler and a standardization effort is already under way in the automotive community to integrate LET into AUTOSAR. We have organized a Dagstuhl seminar [9] to discuss and sketch solutions to the problems raised by the use of LET in multicore systems. A white paper on the topic is under preparation.

So far, LET has been applied only at the ECU (Electronic Control Unit) level by the automotive industry. Recent developments in electric powertrains and autonomous vehicle functions raise parallel programming from the multicore level to the vehicle level where the standard LET approach cannot apply directly. We have proposed System Level LET [21], an extension of LET with relaxed synchronization requirements which allows separating network design from ECU design and makes LET applicable to automotive distributed systems.

6.2.7. Scheduling under multiple constraints and Pareto optimization

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 [26]. 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 visited the team in 2018.

Second, we have proposed the first of its kind multi-criteria scheduling heuristics for a DAG of tasks onto an homogeneous multi-core chip. Given an application modeled as a Directed Acyclic Graph (DAG) of tasks and a multicore architecture, we produce a set of non-dominated (in the Pareto sense) static schedules of this DAG onto this multicore. The criteria we address are the execution time, reliability, power consumption, and peak temperature. These criteria exhibit complex antagonistic relations, which make the problem challenging. For instance, improving the reliability requires adding some redundancy in the schedule, which penalizes the execution time. To produce Pareto fronts in this 4-dimension space, we transform three of the four criteria into constraints (the reliability, the power consumption, and the peak temperature), and we minimize the fourth one (the execution time of the schedule) under these three constraints. By varying the thresholds used for the three constraints, we are able to produce a Pareto front of non-dominated solutions. Each Pareto optimum is a static schedule of the DAG onto the multicore. We propose two algorithms to compute static schedules. The first is a ready list scheduling heuristic called ERPOT (Execution time, Reliability, POwer consumption and Temperature). ERPOT actively replicates the tasks to increase the reliability, uses Dynamic Voltage and Frequency Scaling to decrease the power consumption, and inserts cooling times to control the peak temperature. The second algorithm uses an Integer Linear Programming (ILP) program to compute an optimal schedule. However, because our multi-criteria scheduling problem is NP-complete, the ILP algorithm is limited to very small problem instances. Comparisons showed that the schedules produced by ERPOT are on average only 10% worse than the optimal schedules computed by the ILP program, and that ERPOT outperforms the PowerPerf-PET heuristic from the literature on average by 33%. This is a joint work with Athena Abdi and Hamid Zarandi from Amirkabir University in Tehran, Iran.

6.3. Fault Management and Causal Analysis

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 [48] 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].

We are currently working on a revised version of our general semantic framework for fault ascription in [46] that satisfies a set of formally stated requirements — such as its behavior under several notions of abstraction and refinement —, 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. Fault Management in Virtualized Networks

From a more applied point of view we are investigating, in the context of Sihem Cherrared's PhD thesis, approaches for fault explanation and localization in virtualized networks. In essence, Network Function Virtualization (NFV), widely adopted by the industry and the standardization bodies, is about running network functions as software workloads on commodity hardware to optimize deployment costs and simplify the life-cycle management of network functions. However, it introduces new fault management challenges including dynamic topology and multi-tenant fault isolation that we discuss in [14]. As a first step to tackle those challenges, we have extended the classical fault management process to the virtualized functions by introducing LUMEN: a Global Fault Management Framework. Our approach aims at providing the availability and reliability of the virtualized 5G end-to-end service chain. LUMEN includes the canonical steps of the fault management process and proposes a monitoring solution for all types of Network virtualization Environments. Our framework is based on open source solutions and could easily be integrated with other existing autonomic management models.

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 scheduling parameters for real-time systems modeled as synchronous data flow (SDF), cyclo-static dataflow (CSDF), and ultimately cyclo-static dataflow (UCSDF) graphs. This synthesis aims for a trade-off between throughput maximization and total buffer size minimization. The synthesizer inputs are a 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 synthesized scheduling parameters such as tasks periods, offsets, processor bindings, priorities, buffer initial markings and buffer sizes. In this section, we present new results on two aspects: (1) the improvement of ADFG's usability and tool interoperability, (2) the integration of new scheduling analysis and scheduler synthesis algorithms.

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 relevant metrics to analyze the scheduling efficiency.

Actor models and scheduling algorithms in ADFG are extended to investigate the contention-aware scheduling problem on multi/many-core architectures. The problem we tackled is that the scheduler synthesis for these platforms must account for the non-negligible delay due to shared memory accesses. We exploited the deterministic communications exposed in SDF graphs to account for the contention and further optimize the synthesized schedule. Two solutions are proposed and implemented in ADFG: contention-aware and contention-free scheduling synthesis. In other words, we either take into account the contention and synthesize a contention-aware schedule or find a one that results in no contention.

ADFG is extended to apply a transformation known as partial expansion graphs (PEG). This transformation can be applied as a pre-processing stage to improve the exploitation of data parallelism in SDF graphs on parallel platforms. In contrast to the classical approaches of transforming SDF graphs into equivalent homogeneous forms, which could lead to an exponential increase in the number of actors and excessive communication overhead, PEG-based approaches allow the designer to control the degree to which each actor is expanded. A PEG algorithm that employs cyclo-static data flow techniques is developed in ADFG. Compared to exist PEG-based approach, our solution requires neither buffer managers nor split-join actors to coordinate data production and consumption rates. This allows us to reduce the number of added actors and communication overhead in the expanded graphs.

7.2. Hardware synthesis in Polychrony

Participants: Loïc Besnard, Hafiz Muhamad Amjad.

[0] Real-Time Scheduling of Dataflow Graphs. A. Bouakaz. Ph.D. Thesis, University of Rennes 1, 2013.

[0] The Cheddar project: a GPL real-time scheduling analyzer: http://beru.univ-brest.fr/~singhoff/cheddar/

In the context of the Convex associate-project with the Chinese Academy of Science, we have this year developped code generators (VHDL, Verilog) for modeling hardware in Signal language [0]. The first scheme of the translation had been proposed by Mohammed Belhadj PhD Thesis. Independent on the HDL used, VHDL or Verilog, the translation of Signal to a HDL is quite simple, considering only the functional (executable) Signal programs and a behavioral translation. Indeed, behavioral translation is quite similar to a sequential code generator. In this case, the Signal compiler generates the clock of each SIGNAL signal and orders the execution of the equations. This control structure can be easily translated in the HDL. The generated code may contain conditionals, loops and signal assignements in a HDL process.

7.3. 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 ($d\mathcal{L}$) held our attention [0]. This formalism is built upon the Dynamic Logic of V. Pratt and augmented with the possibility of expressing Ordinary Differential Equations (ODEs). Combined with the ability of Dynamic Logic to specify and verify hybrid programs, $d\mathcal{L}$ is a particularly fit model cyber-physical systems. The proof system associated with the logic is implemented into the theorem prover KeYmaera X. Aimed toward automation, 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[6]. Commutativity provides compositionality: the possibility to compose and prove components and modules in every possible order. Associativity is mandatory to modularly design a system; it allows to construct step-by-step a system by adding new components. We have proved a theorem to automatically build contracts from the composition of components. We have exemplified our results with a cruise-controller example.

This contribution to $d\mathcal{L}$ defines a component-based approach to modularly model and prove cyber-physical systems. We have developed a working prototype in the interactive theorem prover KeYmaera X to show the feasibility of an implementation of our approach. To validate our methodology, we have case studied the example of a water-recycling plant, which rose several challenges.

The timing aspects in cyber-physical systems are a key aspect. A monitor regulating a plant, for example the water-level in a tank, must execute sufficiently often to ensure the correct behavior, for example that the water-level does not overflow. We have adapted our approach to automatically handle the compliance of execution time of monitor with the controllability of plant. We retain commutativity and associativity, thus the modularity of our approach. More importantly, we are still able to automatically build contracts from the composition of components.

We have also adapted our component-based approach to handle modes, a frequent construct in cyber-physical systems. It is also frequent to have to model causal composition between two components, for example that the sensor must execute before the monitor using the data of the sensor. Once again, we have adapted our component-based approach to take into account such design choice. It is important to emphasize that all these adaptations remain within our framework and are thus compatible.

To conclude, we have presented a methodology to tackle complexity of modeling and verification of cyber-physical systems by breaking a systems into smaller parts, the components. We have showed that it is easily adaptable to take into account new challenges. A future work would be to blend our component-based approach with refinement reasoning.

[0] Verilog Code Generation Scheme from Signal Language. Hafiz Muhammad, Amjad and Jianwei, Niu and Kai, Hu and Naveed, Akram and Loïc, Besnard. International Bhurban Conference on Applied Sciences and Technology (IBCAST). IEEE, 2019

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

7.4. Verified information flow of embedded programs

Participants: Jean-Joseph Marty, Jean-Pierre Talpin, Shravan Narayan, Deian Stefan, Rajesh Gupta.

This PhD project is about applying refinement types theory to verified programming of applications and modules of library operting systems, such as unikernels, for embedded devices (the Internet of Things (IoT). We focus on developping a model of information flow control approach using labelled input-outputs (LIO).

We are collaborating with the ProgSys group at UC San Diego in the frame of Inria associate-team Composite, which develops the LIO framework. The LIO framework allows to avoid the "label creep" problem and supports the modeling of concurrency.

Currentlymost of the properties implemented in LIO rely on Haskell properties which is not friendly for embedded devices (IoT), as Haskell requires a huge run-time compared to low resources micro-controllers with less than 32KB of memory.

Instead, we actively use the new Microsoft's verified programming language F*. This programming language is a proof assistant like language that allows us to formalize, verify (using SMT solver and tactics) and extract to clean C (without system dependency) . We succeeded in making proved programs on Arduino compatible micro-controller. Our aim is to develop a version of LIO that could be verified and then extracted to C for targeting operating systems or IoT.

At present, F* is a mix of three domain specific languages: Meta* for proof automation, Low* for system level code including memory safety and F* that glues everything. We successfully implemented a simple Low*-only LIO library allowing to use labeled values. We are now working on a formalized version that will ensure that an F* program is safe w.r.t. information flow, before code generation.

In parallel we continue to work with the ProgSys team on a second project: code-named Gluco*. The goal of this project is to strengthen the F* programming knowledge and to make a example of a safety-critical application where F* can be used [0].

7.5. Modeling cyber-physical systems: from Signal to Signal+

Participants: Thierry Gautier, Albert Benveniste.

Based, initially, on two small case studies, we started a reflection on the modeling and analysis of cyber-physical systems by extending the model of synchronous languages, and in particular that of the Signal language [15]. The principle considered here is to remain within the traditional framework, in discrete time, of synchronous languages, but to completely generalize the equational style of specifications. In the usual Signal language, clocks are defined in a relational way by constraint systems. In contrast, data are defined using functionally inspired dataflow expressions. In this new Signal+, we propose to generalize the equational style to the data: numerical quantities also become subject to constraints. This typically corresponds to the way of modeling a system that includes physical components: for example, equations respecting balance laws have to be written. A major interest is that this programming style is much more compositional than more traditional Simulink or Lustre, or even Signal-based programming. The question then arises as to whether such a program should be analyzed. There is no notion of syntactic dependence, but the incidence graph can be used to define a matching that uniquely associates an equation with each variable. A scheduling can then be synthesized, provided that the reasoning is valid, which is the case for some classes of numerical algebraic equations, for which a solver can be used. However, we can observe on our case studies that the transition to Signal+ class is a real step forward in terms of difficulty. In discrete time, at each reaction, the free variables of the system must be evaluated using what is known about states and inputs. But in some cases, there will be more variables than usable equations. By reasoning on the fact that we have systems that are invariant in time, it may then be necessary to “shift” equations, just as if we were in continuous time, we could differentiate a constraint (a program is a discrete approximation of a continuous time system). This amounts, in a way, to proposing some modifications, which are considered as legitimate, to the source program.

[0] Towards verified programming of embedded devices. J.-P. Talpin, J.-J. Marty, S. Narayan, D. Stefan, R. Gupta. Design, Automation and Test in Europe (DATE'19). IEEE, to appear 2018.

ANTIQUE Project-Team

6. New Results

6.1. A Theoretical Foundation of Sensitivity in an Abstract Interpretation Framework

Participants: Xavier Rival [correspondant], Sukyoung Ryu, Se-Won Kim.

In [14], we formalize a framework to design static analyses that make use of sensitivity, using the general notion of cardinal power abstraction.

Program analyses often utilize various forms of *sensitivity* such as context sensitivity, call-site sensitivity, and object sensitivity. These techniques all allow for more precise program analyses, that are able to compute more precise program invariants, and to verify stronger properties. Despite the fact that sensitivity techniques are now part of the standard toolkit of static analyses designers and implementers, no comprehensive frameworks allow the description of all common forms of sensitivity. As a consequence, the soundness proofs of static analysis tools involving sensitivity often rely on *ad hoc* formalization, which are not always carried out in an abstract interpretation framework. Moreover, this also means that opportunities to identify similarities between analysis techniques to better improve abstractions or to tune static analysis tools can easily be missed.

In this work, we formalize a framework for the description of *sensitivity in static analysis*. Our framework is based on a powerful abstract domain construction, and utilizes reduced cardinal power to tie basic abstract predicates to the properties analyses are sensitive to. We formalize this abstraction, and the main abstract operations that are needed to turn it into a generic abstract domain construction. We demonstrate that our approach can allow for a more precise description of program states, and that it can also describe a large set of sensitivity techniques, both when sensitivity criteria are static (known before the analysis) or dynamic (inferred as part of the analysis), and sensitive analysis tuning parameters. Last, we show that sensitivity techniques used in state of the art static analysis tools can be described in our framework.

6.2. Memory Abstraction

6.2.1. Abstraction of arrays based on non contiguous partitions

Participants: Jiangchao Liu, Xavier Rival [correspondant].

In [15], we studied the verification of components of embedded programs that utilize arrays to store dynamically chained data-structures. Furthermore, this work constitutes a significant part of Jiangchao Liu's PhD Thesis ([10]).

User-space programs rely on memory allocation primitives when they need to construct dynamic structures such as lists or trees. However, low-level OS kernel services and embedded device drivers typically avoid resorting to an external memory allocator in such cases, and store structure elements in contiguous arrays instead. This programming pattern leads to very complex code, based on data-structures that can be viewed and accessed either as arrays or as chained dynamic structures. The code correctness then depends on intricate invariants mixing both aspects. We propose a static analysis that is able to verify such programs. It relies on the combination of abstractions of the allocator array and of the dynamic structures built inside it. This approach allows to integrate program reasoning steps inherent in the array and in the chained structure into a single abstract interpretation. We report on the successful verification of several embedded OS kernel services and drivers.

6.2.2. Semantic-Directed Clumping of Disjunctive Abstract States

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

In [29], we studied the semantic directed clumping of disjunctive abstract states. Furthermore, this work constitutes a significant part of Huisong Li's PhD Thesis ([9]).

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.

6.3. Static Analysis of JavaScript Code

6.3.1. Weakly Sensitive Analysis for Unbounded Iteration over JavaScript Objects

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

In [28], 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.

6.4. Communication-closed asynchronous protocols

Participants: Andrei Damien, Cezara Drăgoi [correspondant], Alexandru Militaru, Josef Widder.

Fault-tolerant distributed systems are implemented over asynchronous networks, so that they use algorithms for asynchronous models with faults. Due to asynchronous communication and the occurrence of faults (e.g., process crashes or the network dropping messages) the implementations are hard to understand and analyze. In contrast, synchronous computation models simplify design and reasoning. In this paper, we bridge the gap between these two worlds. For a class of asynchronous protocols, we introduce a procedure that, given an asynchronous protocol, soundly computes its round-based synchronous counterpart. This class is defined by properties of the sequential code. We computed the synchronous counterpart of known consensus and leader election protocols, such as, Paxos, and Chandra and Toueg's consensus. Using Verifast we checked the sequential properties required by the rewriting. We verified the round-based synchronous counter-part of Multi-Paxos, and other algorithms, using existing deductive verification methods for synchronous protocols.

6.5. Borel Kernels and their Approximation, Categorically

Participants: Fredrik Dahlqvist, Alexandra Silva, Vicent Danos [correspondant], Ilias Garnier.

In [12] is introduced a categorical framework to study the exact and approximate semantics of probabilistic programs. We construct a dagger symmetric monoidal category of Borel kernels where the dagger-structure is given by Bayesian inversion. We show functorial bridges between this category and categories of Banach lattices which formalize the move from kernel-based semantics to predicate transformer (backward) or state transformer (forward) semantics. These bridges are related by natural transformations, and we show in particular that the Radon-Nikodym and Riesz representation theorems-two pillars of probability theory-define natural transformations. With the mathematical infrastructure in place, we present a generic and endogenous approach to approximating kernels on standard Borel spaces which exploits the involutive structure of our category of kernels. The approximation can be formulated in several equivalent ways by using the func-torial bridges and natural transformations described above. Finally, we show that for sensible discretization schemes, every Borel kernel can be approximated by kernels on finite spaces, and that these approximations converge for a natural choice of topology. We illustrate the theory by showing two examples of how approximation can effectively be used in practice: Bayesian inference and the Kleene * operation of ProbNetKAT.

6.6. Static analysis of rule-based models

Thanks to rule-based modeling languages, we can assemble large sets of mechanistic protein-protein interactions within integrated models. Our goal would be to understand how the behavior of these systems emerges from these low-level interactions. Yet this is a quite long term challenge and it is desirable to offer intermediary levels of abstraction, so as to get a better understanding of the models and to increase our confidence within our mechanistic assumptions. To this extend, static analysis can be used to derive various abstractions of the semantics, each of them offering new perspectives on the models.

6.6.1. Trace approximation

Participants: Jérôme Feret [correspondant], Kim Quyên Lý.

In [13], we propose an abstract interpretation of the behavior of each protein, in isolation. Given a model written in Kappa, this abstraction computes for each kind of proteins a transition system that describes which conformations this protein may take and how a protein may pass from one conformation to another one. Then, we use simplicial complexes to abstract away the interleaving order of the transformations between conformations that commute. As a result, we get a compact summary of the potential behavior of each protein of the model.

6.6.2. Detection of polymer formation

Participants: Pierre Boutillier, Aurélie Faure de Pebeyre, Jérôme Feret [correspondant].

Rule-based languages, such as Kappa and BNGL, allow for the description of very combinatorial models of interactions between proteins. A huge (when not infinite) number of different kinds of bio-molecular compounds may arise due to proteins with multiple binding and phosphorylation sites. Knowing beforehand whether a model may involve an infinite number of different kinds of bio-molecular compounds is crucial for the modeler. On the first hand, having an infinite number of kinds of bio-molecular compounds is sometimes a hint for modeling flaws: forgetting to specify the conflicts among binding rules is a common mistake. On the second hand, it impacts the choice of the semantics for the models (among stochastic, differential, hybrid).

In [22], we introduce a data-structure to abstract the potential unbounded polymers that may be formed in a rule-based model. This data-structure is a graph, the nodes and the edges of which are labeled with patterns. By construction, every potentially unbounded polymer is associated to at least one cycle in that graph. This data-structure has two main advantages. Firstly, as opposed to site-graphs, one can reason about cycles without enumerating them (by the means of Tarjan's algorithm for detecting strongly connected components). Secondly, this data-structures may be combined easily with information coming from additional reachability analysis: the edges that are labeled with an overlap that is proved unreachable in the model may be safely discarded.

6.6.3. The static analyzer KaSa

Participants: Pierre Boutillier, Ferdinanda Camporesi, Jean Coquet, Jérôme Feret [correspondant], Kim Quyên Lý, Nathalie Théret, Pierre Vignet.

KaSa is a static analyzer for Kappa models. Its goal is two-fold. Firstly, KaSa assists the modeler by warning about potential issues in the model. Secondly, KaSa may provide useful properties to check that what is implemented is what the modeler has in mind and to provide a quick overview of the model for the people who have not written it. The cornerstone of KaSa is a fix-point engine which detects some patterns that may never occur whatever the evolution of the system may be. From this, many useful information may be collected KaSa warns about rules that may never be applied, about potential irreversible transformations of proteins (that may not be reverted even thanks to an arbitrary number of computation steps) and about the potential formation of unbounded molecular compounds. Lastly, KaSa detects potential influences (activation/inhibition relation) between rules.

In [21], we illustrate the main features of KaSa on a model of the extracellular activation of the transforming growth factor, TGF-b.

6.7. The Kappa platform for rule-based modeling

Participants: Pierre Boutillier, Mutaamba Maasha, Xing Li, Héctor Medina-Abarca, Jean Krivine, Jérôme Feret [correspondant], Ioana Cristescu, Angus Forbes, Walter Fontana.

In [11], we present an overview of the Kappa platform, an integrated suite of analysis and visualization techniques for building and interactively exploring rule-based models. The main components of the platform are the Kappa Simulator, the Kappa Static Analyzer and the Kappa Story Extractor. In addition to these components, we describe the Kappa User Interface, which includes a range of interactive visualization tools for rule-based models needed to make sense of the complexity of biological systems. We argue that, in this approach, modeling is akin to programming and can likewise benefit from an integrated development environment. Our platform is a step in this direction.

We discuss details about the computation and rendering of static, dynamic, and causal views of a model, which include the contact map (CM), snapshots at different resolutions, the dynamic influence network (DIN) and causal compression. We provide use cases illustrating how these concepts generate insight. Specifically, we show how the CM and snapshots provide information about systems capable of polymerization, such as Wnt signaling. A well-understood model of the KaiABC oscillator, translated into Kappa from the literature, is deployed to demonstrate the DIN and its use in understanding systems dynamics. Finally, we discuss how pathways might be discovered or recovered from a rule-based model by means of causal compression, as exemplified for early events in EGF signaling.

The Kappa platform is available via the project website at kappa-language.org. All components of the platform are open source and freely available through the authors' code repositories.

6.8. Conservative approximation of systems of differential equations

We design a tools-kit to reason and abstract the solutions of the systems of differential equations that are described in high-level languages. Our abstractions are conservative in the sense that they provided sound lower and upper bounds for the value of some observables of the system. Our approach consists, firstly, in inferring structural equalities about combinations of variables and structural inequalities about the value of variable derivatives thanks to symbolic reasoning at the level of the languages and, then, in using these numerical constraints to infer two differential equations for the variables of interest — one for the lower bound and one for the upper bound.

We focus on the systems of equations that are described in Kappa. Our goal is to provide a unifying framework that can deal with heterogeneous kinds of abstractions, including truncation, time- and concentration-scale separations, flow-based reduction, symmetries-based reduction.

6.8.1. Approximation of models of polymers

Participants: Ken Chanseau Saint-Germain, Jérôme Feret [correspondant].

We propose a systematic approach to approximate the behavior of models of polymers synthesis/degradation, described in Kappa. Our abstraction consists in focusing on the behavior of all the patterns of size less than a given parameter. We infer symbolic equalities and inequalities which intentionally may be understood as algebraic constructions over patterns, and extensionally as sound properties about the concentration of the bio-molecular species that contain these patterns. Then, we derive a system of equations describing the time evolution of a lower and an upper bounds for the concentration of each pattern of interest.

This work has been presented at VEMDP 2018 (Verification of Engineered Molecular Devices and Programs), in Oxford, 19th July 2018, and at the days “BIOS-IA” of the working group BIOSS, at Pasteur Institute, Paris, 18th December 2018.

6.8.2. Approximation based on time- and/or concentration-scale separation

Participants: Andreea Beica, Jérôme Feret [correspondant].

In [20], we have designed and tested an approximation method for ODE models of biochemical reaction systems, in which the guarantees are our major requirement. Borrowing from tropical analysis techniques, we look at the dominance relations among terms of each species’ ODE. These dominance relations can be exploited to simplify the original model, by neglecting the dominated terms. As the dominant subsystems can change during the system’s dynamics, depending on which species dominate the others, several possible modes exist. Thus, simpler models consisting of only the dominant subsystems can be assembled into hybrid, piece-wise smooth models, which approximate the behavior of the initial system. By combining the detection of dominated terms with symbolic bounds propagation, we show how to approximate the original model by an assembly of simpler models, consisting in ordinary differential equations that provide time-dependent lower and upper bounds for the concentrations of the initial models species. The utility of our method is twofold. On the one hand, it provides a reduction heuristics that performs without any prior knowledge of the initial system’s behavior (i.e., no simulation of the initial system is needed in order to reduce it). On the other hand, our method provides sound interval bounds for each species, and hence can serve to evaluate the faithfulness of tropicalization reduction heuristics for ODE models of biochemical reduction systems. The method is tested on several case studies.

6.9. Sources, propagation and consequences of stochasticity in cellular growth

Participants: Philipp Thomas, Guillaume Terradot, Vicent Danos [correspondant], Andrea Weiße.

Growth impacts a range of phenotypic responses. Identifying the sources of growth variation and their propagation across the cellular machinery can thus unravel mechanisms that underpin cell decisions.

In [17], we present a stochastic cell model linking gene expression, metabolism and replication to predict growth dynamics in single bacterial cells. Alongside we provide a theory to analyze stochastic chemical reactions coupled with cell divisions, enabling efficient parameter estimation, sensitivity analysis and hypothesis testing. The cell model recovers population-averaged data on growth-dependence of bacterial physiology and how growth variations in single cells change across conditions. We identify processes responsible for this variation and reconstruct the propagation of initial fluctuations to growth and other processes. Finally, we study drug-nutrient interactions and find that antibiotics can both enhance and suppress growth heterogeneity. Our results provide a predictive framework to integrate heterogeneous data and draw testable predictions with implications for antibiotic tolerance, evolutionary and synthetic biology.

6.10. Survival of the Fattest: Evolutionary Trade-offs in Cellular Resource Storage

Participants: Guillaume Terradot, Andreea Beica, Andrea Weiße, Vicent Danos [correspondant].

Cells derive resources from their environments and use them to fuel the bio-synthetic processes that determine cell growth. Depending on how responsive the bio-synthetic processes are to the availability of intracellular resources, cells can build up different levels of resource storage.

In [16], we use a recent mathematical model of the coarse-grained mechanisms that drive cellular growth to investigate the effects of cellular resource storage on growth. We show that, on the one hand, there is a cost associated with high levels of storage resulting from the loss of stored resources due to dilution. We further show that, on the other hand, high levels of storage can benefit cells in variable environments by increasing biomass production during transitions from one medium to another. Our results thus suggest that cells may face trade-offs in their maintenance of resource storage based on the frequency of environmental change.

6.11. A Genetic Circuit Compiler: Generating Combinatorial Genetic Circuits with Web Semantics and Inference

Participants: William Waites, Goksel Misirli, Matteo Cavaliere, Vicent Danos [correspondant].

A central strategy of synthetic biology is to understand the basic processes of living creatures through engineering organisms using the same building blocks. Biological machines described in terms of parts can be studied by computer simulation in any of several languages or robotically assembled in vitro. In [19] we present a language, the Genetic Circuit Description Language (GCDL) and a compiler, the Genetic Circuit Compiler (GCC). This language describes genetic circuits at a level of granularity appropriate both for automated assembly in the laboratory and deriving simulation code. The GCDL follows Semantic Web practice and the compiler makes novel use of the logical inference facilities that are therefore available. We present the GCDL and compiler structure as a study of a tool for generating κ-language simulations from semantic descriptions of genetic circuits.

6.12. An Information-Theoretic Measure for Patterning in Epithelial Tissues

Participants: William Waites, Matteo Cavaliere, Élise Cachat, Vicent Danos [correspondant], Jamie A. Davies.

In [18], we present path entropy, an information-theoretic measure that captures the notion of patterning due to phase separation in organic tissues. Recent work has demonstrated, both in silico and in vitro, that phase separation in epithelia can arise simply from the forces at play between cells with differing mechanical properties. These qualitative results give rise to numerous questions about how the degree of patterning relates to model parameters or underlying biophysical properties. Answering these questions requires a consistent and meaningful way of quantifying degree of patterning that we observe. We define a resolution-independent measure that is better suited than image-processing techniques for comparing cellular structures. We show how this measure can be usefully applied in a selection of scenarios from biological experiment and computer simulation, and argue for the establishment of a tissue-graph library to assist with parameter estimation for synthetic morphology.

CELTIQUE Project-Team

5. New Results

5.1. Software Fault Isolation

Participants: Frédéric Besson, Thomas Jensen, Julien Lepiller.

Software Fault Isolation (SFI) consists in transforming untrusted code so that it runs within a specific address space, (called the sandbox) and verifying at load-time that the binary code does indeed stay inside the sandbox. Security is guaranteed solely by the SFI verifier whose correctness therefore becomes crucial. Existing verifiers enforce a very rigid, almost syntactic policy where every memory access and every control-flow transfer must be preceded by a sandboxing instruction sequence, and where calls outside the sandbox must implement a sophisticated protocol based on a shadow stack. We have defined SFI as a defensive semantics, with the purpose of deriving semantically sound verifiers that admit flexible and efficient implementations of SFI. We derive an executable analyser, that works on a per-function basis, which ensures that the defensive semantics does not go wrong, and hence that the code is well isolated. Experiments show that our analyser exhibits the desired flexibility: it validates correctly sandboxed code, it catches code breaking the SFI policy, and it can validate programs where redundant instrumentations are optimised away [8].

5.2. Compilation and Side-Channels

Participants: Frédéric Besson, Alexandre Dang, Thomas Jensen.

The usual guarantee provided by compilers is that the input/output observable behaviour of the target program is one of the possible behaviours of the source program. In the context of security, the notion of observable behaviour needs to be revisited in order to take into account side-channels *i.e.* observations beyond input/output that are leaked by the program execution to a potential attacker.

For instance, a common security recommendation is to reduce the in-memory lifetime of secret values, in order to reduce the risk that an attacker can obtain secret data by probing memory. To mitigate this risk, secret values can be overwritten, at source level, after their last use. However, as secret values are never used afterwards, a compiler may remove this mitigation during a standard Dead Store Elimination pass. We propose a formal definition of Information Flow Preserving transformation [7] which ensures that secret values are not easier to obtain at assembly level than at source level. Using the notion of Attacker Knowledge, we relate the information leak of a program before and after the transformation. We consider two classic compiler passes (Dead Store Elimination and Register Allocation) and show how to validate and, if needed, modify these transformations in order to be information flow preserving.

5.3. Semantic study of the Sea-of-Nodes form

Participants: Delphine Demange, Yon Fernandez de Retana, David Pichardie.

As part of the PhD of Yon Fernandez de Retana [2], we started to study the the Sea-of-Nodes form. This intermediate representation was introduced by Cliff Click in the mid 90s [21] as an enhanced SSA form. It improves on the initial SSA form by relaxing the total order on instructions in basic blocks into explicit data and control dependencies. This makes programs more flexible to optimize. While Sea-of-node is popular in many production-size compiler (Sun's HotSpot, Graal...), it is still not very well understood, from a semantic, foundational point of view. We have defined a simple but rigorous formal semantics for a Sea-of-Nodes form. It comprises a denotational component to express data computation, and an operational component to express control flow. We prove a fundamental, dominance-based semantic property on Sea-of-Nodes programs which determines the regions of the graph where the values of nodes are preserved. Finally, we apply our results to prove the semantic correctness of a redundant zero-check elimination optimization. All the necessary semantic properties have been mechanically verified in the Coq proof assistant. These results have been published in [11]. A more detailed account can be found in Yon Fernandez de Retana's PhD manuscript [2].

5.4. Certified Concurrent Garbage Collector

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

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 [6] as an extended version of [22].

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 [20].

5.5. Formalization of Higher-Order Process Calculi

Participants: Guillaume Ambal, Sergueï Lenglet, Alan Schmitt.

Guillaume Amabal, Sergueï Lenglet, Alan Schmitt have continued exploring how to formalize HOπ in Coq, in particular how to deal with the different kinds of binders used in the calculus. We have studied and compared several approaches, such as locally nameless, De Bruijn indices, and nominal binders. We have discovered that the locally nameless approach introduces a lot of complexity, as name restriction allows reduction under binders, introducing the need for numerous renaming lemmas. The nominal approach is quite elegant and very close to the pen and paper definitions, but it still requires many technical lemmas to be proven. The de Bruijn approach is the most concise. A first version of this work has been published at CPP 2018 [18] and a journal version is submitted for publication. The Coq scripts can be found at http://passivation.gforge.inria.fr/hopi/.

5.6. Certified Semantics and Analyses for JavaScript

Participants: Samuel Risbourg, 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. Since September 2018, Samuel Risbourg has been hired to continue developing the tool. It is publicly available at https://github.com/jscert/jsexplain and is described in a publication at The Web Conference 2018 [10]. The tool is regularly presented at the TC39 committee standardizing JavaScript to solicit feedback.

5.7. Skeletal Semantics

Participants: Nathanael Courant, Thomas Jensen, Alan Schmitt.

Alan Schmitt and Thomas Jensen, in collaboration with Martin Bodin and Philippa Gardner at Imperial College London, have designed a new meta-language to formally describe semantics. A fundamental idea behind this approach to semantics is that this description can be used to derive several *interpretations* corresponding to different kinds of semantics, such as a big-step semantics, an abstract interpretation, or a control-flow analysis. The correctness of these semantics is proven independently of the language considered. This work has been accepted at POPL 2019 [4] and is formalized in Coq (see the skeletal semantics web site for more detail). Nathanael Courant is currently extending this work to generate analyses automatically and to facilitate the way in which to adjust their precision.

5.8. Verification of High-Level Transformations with Inductive Refinement Types

Participant: Thomas Jensen.

High-level transformation languages like Rascal or Stratego include expressive features for manipulating large abstract syntax trees: first-class traversals, expressive pattern matching, backtracking and generalized iterators. We have designed and implemented an abstract interpretation tool, Rabit, for verifying inductive type and shape properties for transformations written in such languages. We describe how to perform abstract interpretation based on operational semantics, specifically focusing on the challenges arising when analyzing the expressive traversals and pattern matching. We have evaluated Rabit on a series of transformations (normalization, desugaring, refactoring, code generators, type inference, etc.) showing that we can effectively verify stated properties.

This work was done in collaboration with researchers at the IT University of Copenhagen. The paper [19] presenting these results won the Best Paper award at GPCE 2018.

5.9. 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. This year we successfully extended this result to the case of higher-order functional programs [15], [16]. Moreover, the approximation automaton can be certified using an efficient Coq-extracted checker that we developped in 2008. Thus, we have an automatic static analysis procedure for higher-order functional programs whose results are certified by the Coq proof assistant. The algorithm presented in [15] has been implemented in Timbuk [14] and gives very encouraging experimental results http://people.irisa.fr/Thomas.Genet/timbuk/funExperiments/. Besides, we have shown the completeness of this approach, i.e., that any regular approximation of the image of a function can be found using completion [13].

CONVECS Project-Team

6. New Results

6.1. New Formal Languages and their Implementations

6.1.1. LOTOS and LNT Specification Languages

Participants: Hubert Garavel, Frédéric Lang, Wendelin Serwe.

LNT [5] [36] is a next-generation formal description language for asynchronous concurrent systems. The design of LNT at CONVECS is the continuation of the efforts undertaken in the 80s to define sound languages for concurrency theory and, indeed, LNT is derived from the ISO standards LOTOS (1989) and E-LOTOS (2001). In a nutshell, LNT attempts to combine the best features of imperative programming languages, functional languages, and value-passing process calculi.

LNT is not a frozen language: its definition started in 2005, as part of an industrial project. Since 2010, LNT has been systematically used by CONVECS for numerous case studies (many of which being industrial applications — see § 6.5). LNT is also used as a back-end by other research teams who implement various languages by translation to LNT. It is taught in university courses, e.g., at University Grenoble Alpes and ENSIMAG, where it is positively accepted by students and industry engineers. Based on the feedback acquired by CONVECS, LNT is continuously improved.

In 2018, the CADP tools that translate LNT to LOTOS have been enhanced in various ways. In the warning and error messages emitted by LNT2LOTOS, line numbers have been made more precise. In addition to a bug fix, the LNT_DEPEND tool, which computes dependencies between LNT modules has been entirely rewritten and made much faster. Also, the LNT language has been simplified by removing “!external” pragmas for constructors, as “!external” pragmas for types are sufficient.

We also continued improving the TRAIAN compiler for the LOTOS NT language (a predecessor of LNT), which is used for the construction of most CADP compilers and translators.

In February 2018, we released version 2.9 of TRAIAN. We scrutinized the source code of TRAIAN, deleting all parts of code corresponding to those features of the LOTOS NT language that were either not fully implemented or seldom used in practice. This reduced the source code of TRAIAN by 40% and the binaries by 50%. External LOTOS NT functions are now allowed to return a non-void result. Support for 64-bit macOS executables was added. A few bugs have been fixed and the reference manual of TRAIAN was entirely revised.

The main limitation of TRAIAN 2.x is that it is a 20-year-old compiler that is increasingly difficult to maintain. It consists in a large collection of attribute grammars and is built using the FNC-2 compiler generation system, which is no longer supported. For this reason, TRAIAN only exists in 32-bit version, and sometimes hits the 3–4 GB RAM limit when dealing with large compiler specifications, such as those of LNT2LOTOS or EVALUATOR 5.

For this reason, we undertook a complete rewrite of TRAIAN to get rid of FNC-2. Two main design decisions behind TRAIAN 3.0 are the following: (i) it supports (most of) the LOTOS NT language currently accepted by TRAIAN 2.9, but also extensions belonging to LNT, so as to allow a future migration from LOTOS NT to LNT; and (ii) TRAIAN 3.0 is currently written in LOTOS NT and compiled using TRAIAN 2.9, but should be ultimately capable of bootstrapping itself.

So far, a lexer and parser for LOTOS NT have been developed using the SYNTAX compiler-generation system [0] developed at Inria Paris. This work triggered an in-depth reexamination of the programming interfaces offered by SYNTAX and led to enhancements of these interfaces (see § 6.1.6).

[0] http://syntax.gforge.inria.fr

The abstract syntax tree of LOTOS NT, and the library of predefined LOTOS NT types and functions have been redesigned; previously specified as FNC-2 attribute grammars, they are now themselves written in LOTOS NT, so as to allow bootstrap, using the current version of TRAIAN to build the next one. The construction of the abstract syntax tree has also been completed. Finally, we set several non-regression test bases gathered all available programs written in LOTOS NT.

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 [39] and its PNML representation is described here [0].

The NUPN model has been adopted by the Model Checking Contest and the Rigorous Examination of Reactive Systems challenge. It has been so far implemented in thirteen different tools developed in four countries.

In 2018, a journal article (to appear in 2019) has been written to formalize the complete theory of NUPNs. The CAESAR.BDD tool for NUPNs has been extended with twelve new options. A new tool named NUPN_INFO has been added to CADP to perform three normalizing transformations of NUPNs.

6.1.3. MCL and XTL Property Specification Languages

Participants: Hubert Garavel, Radu Mateescu.

CADP provides two different languages, named MCL and XTL, for expressing data-handling temporal properties of concurrent systems. MCL is an extension of alternation-free modal μ-calculus with data values, programming language constructs, generalized regular formulas on transition sequences, and fairness operators. XTL is a functional-like programming language interpreted on Labeled Transition Systems, enabling the definition of temporal operators by computing their interpretation using fixed point iterations over sets of states and transitions.

In 2018, we enhanced these languages and their associated tools as follows:

- The MCL v4 language was enhanced with a new operator "**loop**" on regular formulas over transition sequences. This general iteration operator parameterized by data variables is able to characterize complex (recursively definable) sequences in an LTS. Two auxiliary regular operators "**continue**" and "**exit**" carrying data values were also introduced to express the repetition and the termination of a loop regular formula, respectively. These operators are particularly useful for specifying transition sequences having a particular cumulated cost (e.g., number of transitions, sum of weights associated to actions, etc.) in the context of probabilistic verification (see § 6.3.2).
- The MCL v3 language was modified and aligned on MCL v4 by removing syntactic differences that existed between both languages concerning the infinite repetition operator ("@") and the respective precedences of the concatenation (".") and choice ("|") operators in regular formulas. MCL v3 has also been enriched with the option operator ("?") on regular formulas already present in MCL v4.
- Consequently, the two versions of MCL_EXPAND for MCL v3 and MCL v4 have been unified in one single tool, which is now invoked by both EVALUATOR 3 and EVALUATOR 4. The corresponding manual pages have been simplified accordingly, with the introduction of two overarching manual pages ("mcl" and "evaluator"). In addition to five bug fixes, the memory footprint of MCL_EXPAND has been reduced. The error messages displayed by MCL_EXPAND, EVALUATOR 3, and EVALUATOR 4 have been improved in terms of accuracy and explanatory contents.

[0] http://mcc.lip6.fr/nupn.php

- In addition to four bug fixes, the XTL model checker now performs consistency checks on the C identifiers specified by the pragmas "!implementedby", "!comparedby", "!enumeratedby", and "!printedby".
- Two new options were added to the EVALUATOR and XTL model checkers: "-depend", which displays the libraries transitively included in an MCL or XTL file, and "-source", which is used by SVL to display correct file names and line numbers for MCL or XTL formulas embedded in SVL scenarios.

6.1.4. Translation of Term Rewrite Systems

Participant: Hubert Garavel.

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 2018, we corrected and/or enhanced several of the existing REC translators and finalized experiments. The results of this study have been presented during an invited talk at WRLA'2018 (*12th International Workshop on Rewriting Logic and its Applications*) and an article [15] was published in the WRLA post-proceedings.

6.1.5. Formal Modeling and Analysis of BPMN

Participant: Gwen Salaün.

A business process is a set of structured activities that provide a certain service or product. Business processes can be modeled using the BPMN standard, and several industrial platforms have been developed for supporting their design, modeling, and simulation.

In collaboration with Francisco Durán and Camilo Rocha (University of Málaga, Spain), we proposed a rewriting logic executable specification of BPMN with time and extended with probabilities. Duration times and delays for tasks and flows can be specified as stochastic expressions, while probabilities are associated to various forms of branching behavior in gateways. These quantities enable discrete-event simulation and automatic stochastic verification of properties such as expected processing time, expected synchronization time at merge gateways, and domain-specific quantitative assertions. The mechanization of the stochastic analysis tasks is done with Maude's statistical model checker PVeStA. These results led to a publication in an international journal [10].

We also worked on an extension of BPMN with data, which is convenient for describing real-world processes involving complex behavior and data descriptions. By considering this level of expressiveness due to the new features, challenging questions arise regarding the choice of the semantic framework for specifying such an extension of BPMN, as well as how to carry out the symbolic simulation, validation, and assess the correctness of the process models. These issues were addressed first by providing a symbolic executable rewriting logic semantics of BPMN using the rewriting modulo SMT framework, where the execution is driven by rewriting modulo axioms and by querying SMT decision procedures for data conditions. Second, reachability properties, such as deadlock freedom and detection of unreachable states with data exhibiting certain values, can be specified and automatically checked with the help of Maude, thanks to its support for rewriting modulo SMT. These results led to a publication in an international conference [21].

6.1.6. Other Language Developments

Participants: Hubert Garavel, Frédéric Lang, 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 2018, we enhanced the SYNTAX compiler generator [0] in various ways: (i) The “string manager” has been generalized to allow several symbol tables to be handled simultaneously; (ii) The “source manager” has been extended with new relocation primitives that enable the caller to specify alternative file names and line numbers for the source file being parsed; for instance, this is typically useful for implementing the “#line” pragma of the C preprocessor; this mechanism has been extended to transparently handle multiple relocations (triggered by the lexer) while recognizing the right-hand side of a syntax rule in the grammar; (iii) The “include manager” has been modified to store file names in a distinct symbol table than the table of identifiers, and to provide the list of all files transitively included from the principal module; (iv) Finally, the main programming interface of SYNTAX has been extended with new primitives, so that at present only 5 calls (rather than 9–13 calls, formerly) are required to launch a compiler written using SYNTAX.

All the CADP compilers have been modified to take advantage of the improvements of the SYNTAX library.

Also, a master student started to study an automated translation from Event-B to LNT. He reviewed the syntax and semantics of Event-B and proposed a pencil-paper translation of most Event-B operators. He applied it to a small example consisting of a bank system, where accounts can be created and closed, and money can be deposited or withdrawn. This was a preliminary work that did not lead to a full implementation, due to lack of time. However, this work is a solid basis for a later implementation.

6.2. Parallel and Distributed Verification

6.2.1. Distributed State Space Manipulation

Participant: Wendelin Serwe.

For distributed verification, CADP provides the PBG format, which implements the theoretical concept of *Partitioned LTS* [44] and provides a unified access to an LTS distributed over a set of remote machines.

In 2018, we improved the usability of distributed state space manipulation tools. In particular:

- A memory shortage error that occurs on a computing node now triggers a distributed termination of the computation, producing proper error messages in the log file of that node.
- A similar naming scheme for log files produced by computing nodes was enforced for all distributed verification tools, which prevents interferences between different invocations of the tools.

6.2.2. Debugging of Concurrent Systems using Counterexample Analysis

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, (iii) the counterexample does not explicitly highlight the source of the bug that is hidden in the model, (iv) the most relevant actions are not highlighted in the counterexample, and (v) the counterexample does not give a global view of the problem.

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. This is achieved by detecting in the models some specific choices between transitions leading to a correct behaviour or falling into an erroneous part of the model. These choices, which we call “neighbourhoods”, provide key information for understanding the bug behind the counterexample. To extract such choices, we proposed a first method for debugging the counterexamples of safety property violations. To do so, it builds a new model from the original one containing all the counterexamples, and then compares the two models to identify neighbourhoods.

[0] http://syntax.gforge.inria.fr

In 2018, we proposed a different method for debugging the counterexamples of liveness property violations. Given a liveness property, it extends the model with prefix and suffix information w.r.t. that property. This enriched model is then analysed to identify neighbourhoods. A set of abstraction techniques we developed exploit the enriched model annotated with neighbourhoods to extract relevant actions from counterexamples, which makes their comprehension easier. This work led to a publication in an international conference [16].

Both approaches are fully automated by a tool we implemented and that has been validated on real-world case studies from various application areas. We extended the methodology and tool with 3D visualization techniques to visualize the erroneous part of the model with a specific focus on neighbourhoods, in order to have a global view of the bug behaviour. This work led to a publication to appear in an international conference.

A detailed description of the proposed methodology is available in G. Barbon's PhD thesis [8].

6.3. Timed, Probabilistic, and Stochastic Extensions

6.3.1. Tools for Probabilistic and Stochastic Systems

Participants: Hubert Garavel, Frédéric Lang.

Formal models and tools dealing with quantitative 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 cyber-physical systems. The wealth of such formal models is sometimes referred to as a quantitative "zoo" [48].

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*) [50]. 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 2018, BCG_STEADY and BCG_TRANSIENT were enhanced along the following lines:

- They were extended to handle single-state Markov chains and to properly compute state solution vectors and transition throughputs on such models.
- Their command-line options were simplified and warnings are emitted when the input Markov chain contains no stochastic transition.
- A problem which caused correct Markov chains to be rejected was corrected. This problem was due to floating point conversion and rounding errors.
- A confusion between state numbers and matrix indices was fixed in the output and error messages.
- Models containing probabilistic self-loops are now rejected, as was already the case of longer circuits of probabilistic transitions, as both represent similar "timelock" situations.

6.3.2. On-the-fly Model Checking for Extended Regular Probabilistic Operators

Participant: Radu Mateescu.

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*) [47], 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 2018, we continued improving and using the extended model checker as follows:

- The model checker now determinizes the dataless regular formulas contained in regular probabilistic operators, ensuring automatically that the linear equation systems produced by the verification of these operators have a unique solution.
- For nondeterministic data-handling regular formulas contained in regular probabilistic operators, the model checker now produces a warning message informing the user that the determinization has to be done manually.
- We carried out further experiments to analyze the quantitative behaviour of the Bounded Retransmission Protocol, namely the variation of the probability of transmission failure w.r.t. the total number of retransmissions attempts.

A paper describing the probabilistic extension of MCL and of the on-the-fly model checker was published in an international journal [13].

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 2018, we improved these tools along the following lines:

- SVL now invokes EVALUATOR 3, EVALUATOR 4, and XTL with their new "-source" option, so that error and warning messages regarding temporal logic formulas now display line numbers in the SVL file itself, rather than in the temporary files generated to contain the temporal logic formulas, making it easier for users to modify incorrect MCL and XTL formulas contained in SVL files.
- SVL has been modified so that both EVALUATOR 3 and EVALUATOR 4 can now be used to compute "deadlock" and "livelock" statements.
- SVL does not require anymore that every "property" statement contains at least one verification statement, namely "comparison", "verify", "deadlock", "livelock", or a shell-line command with an "expected" clause.
- In addition to a bug fix, the EXP.OPEN tool was enhanced with a new option "-depend", displaying both the list of EXP files included (directly or transitively) in the input EXP file, and the list of automata, hide, rename, and cut files used (directly or transitively) in the input EXP file.

A paper containing both a tutorial and a survey on compositional verification was published in an international conference [14].

6.4.2. On-the-Fly Test Generation

Participants: Lina Marsso, Radu Mateescu, Wendelin Serwe.

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.

To address these limitations, we developed the prototype tool TESTOR [0] to extract test cases completely on the fly. TESTOR presents several advantages w.r.t. TGV: (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.

In 2018, we improved TESTOR and TGV as follows:

- TESTOR has been ported to the Windows operating system.
- TESTOR can now be directly connected (by means of Unix pipes) to a system under test (SUT), executing the test case, rather than generating an abstract test-case that has to be connected to the SUT.
- We revised the architecture of TESTOR, so that the interface for the user is more similar to the one of TGV. This enables a user to easily switch between both tools.
- Taking advantage of the similar interfaces, we merged the non-regression test bases of TESTOR and TGV.
- We also fixed a bug and added a new option "-self" to TGV, reducing the number of warning messages.

These activities led to a new version 3.0 of TESTOR and two publications in international conferences [24], [18].

6.4.3. Other Component Developments

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

In 2018, several components of CADP have been improved as follows:

- The CADP toolbox now contains a new tool named SCRUTATOR for pruning Labeled Transition Systems on the fly.
- The OPEN/CAESAR environment was enriched with a new SOLVE_2 library for solving linear equation systems on the fly.
- Two manual pages ("bes" and "seq") have been added, which provide standalone definitions of CADP's BES format for Boolean Equation Systems and SEQ format for execution traces. The OPEN/CAESAR manual pages have been enhanced to give full prototypes for function parameters.
- The CADP toolbox has been ported to Solaris 11 and to SunOS 5.11 OpenIndiana "Hipster". CADP has also been ported to macOS 10.14 "Mojave" and a 64-bit version of CADP is now available for macOS.
- We also designed new C functions for handling path names in order to replace the traditional POSIX primitives basename(), dirname(), and realpath(), which suffer from limitations and ambiguities.

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.

Fog computing provides computing, storage and communication resources (and devices) at the edge of the network, near the physical world (PW). These end-devices nearing the physical world can have interesting properties such as short delays, responsiveness, optimized communications and privacy, which are especially appealing to IoT (Internet of Things) applications. However, IoT devices in the fog have low stability and are prone to failures.

[0] http://convecs.inria.fr/software/testor

In the framework of the collaboration with Orange Labs (see § 7.1.1), we are working on the key challenge of providing reliable services. This may be critical in this context since the non-containment of failures may impact the physical world. For instance, the failure of a smoke detector or a lamp in a smart home for elderly/medicated people may be hazardous. The design of such resilience solutions is complex due to the specificities of the environment, i.e., (i) dynamic infrastructure, where entities join and leave without synchronization; (ii) high heterogeneity in terms of functions, communication models, network, processing and storage capabilities; and (iii) cyber-physical interactions, which introduce non-deterministic and physical world's space and time dependent events.

In 2018, our work focused on proposing an end-to-end resilience approach for stateful IoT applications in the fog taking into account the three specificities mentioned above. The resilience protocol is functionally divided into four phases: (i) state-saving; (ii) monitoring and failure detection; (iii) failure notification and reconfiguration; and (iv) decision and recovery. The protocol implements a combination of different state-saving techniques based on rules and policies to cope with the heterogeneous nature of the environment and recover from failures in a consistent way, including PW-consistency. This work led to a publication in an international conference [25].

To illustrate our protocol at work, we mounted a smart home testbed with objects that can be found in real-life smart homes to test our solution. Our resilience approach was also implemented as a framework and deployed onto the testbed. The empirical results showed that multiple failures are recovered in an acceptable time in regard to end users. This work led to a publication to appear in an international conference.

6.5.2. Verified Composition and Deployment of IoT Applications

Participants: Radu Mateescu, Ajay Muroor Nadumane, Gwen Salaün.

The Internet of Things (IoT) is an interconnection of physical devices and software entities that can communicate and perform meaningful tasks largely without human intervention. The design and development of IoT applications is an interesting problem as these applications are typically dynamic, distributed and, more importantly, heterogeneous in nature.

In the framework of the collaboration with Nokia Bell Labs (see § 7.1.2), we proposed to build and deploy reliable IoT applications using a series of steps: (i) IoT objects and compositions are described using an interface-based behavioural model; (ii) the correctness of the composition is ensured by checking a behavioural compatibility notion that we proposed for IoT systems; and (iii) finally, a deployment plan respecting the dependencies between the objects is generated to facilitate automated deployment and execution of the application.

Regarding implementation, behavioural models and composition are specified in LNT and we take advantage of the CADP toolbox to perform compatibility checks. The deployment is automated using the Majord'Home platform developed by Nokia Bell Labs. The entire implementation is packaged as a Web tool available for end-users. This work led to a publication to appear in an international conference.

6.5.3. Memory Protection Unit

Participants: Hubert Garavel, Radu Mateescu, Wendelin Serwe.

Asynchronous circuits have key advantages in terms of low energy consumption, robustness, and security. However, the absence of a global clock makes the design prone to deadlock, livelock, synchronization, and resource-sharing errors. Formal verification is thus essential for designing such circuits, but it is not widespread enough, as many hardware designers are not familiar with it and few verification tools can cope with asynchrony on complex designs. In the framework of the SECURIOT-2 project (see § 8.2.2.1), we are interested in the rigorous design of asynchronous circuits used in the secure elements for IoT devices developed in the project.

In collaboration with Aymane Bouzafour and Marc Renaudin (Tiempo Secure), we suggested an extension of Tiempo's industrial design flow for asynchronous circuits, based upon the standard Hardware Description Language SystemVerilog (SV), with the formal verification capabilities provided by CADP. This was achieved by translating SV descriptions into LNT, expressing correctness properties in MCL, and verifying them using the EVALUATOR model checker of CADP. It turned out that the constructs of SV and LNT are in close correspondence, and that the synthesizable SV subset can be entirely translated into LNT. The MCL language was also shown adequate for expressing all property patterns relevant for asynchronous circuits.

The practicality of the approach was demonstrated on an asynchronous circuit (4000 lines of SV) implementing a memory protection unit (MPU). The MPU block exhibits a high degree of internal concurrency, comprising 660 parallel execution flows and 250 internal communication channels. The corresponding state space was generated compositionally, by identifying a suitable minimization and composition strategy described in SVL (the largest intermediate state space had more than 116 million states and 862 million transitions). A set of 184 MCL properties were successfully verified on the state space, expressing the correct initialization of the MPU configuration registers, the mutual exclusion of read and write operations on registers, the correct responses to stimuli, and the security requirements related to the many access-control policies enforced by the MPU. This work led to a publication in an international conference [17].

6.5.4. TLS 1.3 Handshake Protocol

Participants: Lina Marsso, Radu Mateescu.

Security services are extensively used in fields like online banking, e-government, online shopping, etc. To ensure a secure communication between peers in terms of authenticity, privacy, and data integrity, cryptographic protocols are applied to regulate the data transfer. These protocols provide a standardized set of rules and methods for the interaction between peers. The Transport Layer Security (TLS) is a widely used security protocol, encompassing a set of rules for the communication between clients and servers, and relying on public-key cryptography to ensure integrity of exchanged data. However, despite multiple prevention measurements, several vulnerabilities (such as Heartbleed and DROWN), have been discovered recently. Therefore, testing the implementations of security protocols is still a crucial issue.

In the framework of the RIDINGS PHC project (see § 8.3.1), we are interested in testing protocols and distributed systems. In collaboration with Josip Bozic and Franz Wotawa (TU Graz, Austria), we undertook the formal modelling of the draft TLS 1.3 handshake protocol [0]. Taking as input the informal description of TLS 1.3 in the draft standard, we developed a formal model (1293 lines of LNT) specifying the handshake messages and client-server interactions. As far as we are aware, this is the first formal model of the draft TLS 1.3 handshake.

We used our LNT model for conformance testing with the OpenSSL version 1.0.1e implementation of the TLS protocol [0]. We defined three test purposes specifying requirements from the draft TLS 1.3 handshake, and applied the newly developed TESTOR tool (see § 6.4.2) to generate the test cases from the LNT model and each test purpose. The execution of these test cases on the OpenSSL implementation spotted a discrepancy of the server's response to a client certificate request w.r.t. the draft TLS 1.3 standard. This work led to a publication in an international workshop [18].

6.5.5. Message Authenticator Algorithm

Participants: Hubert Garavel, Lina Marsso.

The Message Authenticator Algorithm (MAA) is one of the first cryptographic functions for computing a Message Authentication Code. Between 1987 and 2001, the MAA was adopted in international standards (ISO 8730 and ISO 8731-2) to ensure the authenticity and integrity of banking transactions. The MAA also played a role in the history of formal methods, as the National Physical Laboratory (NPL, United Kingdom) developed, in the early 90s, three formal, yet non-executable, specifications of the MAA in VDM, Z, and LOTOS abstract data types.

[0] https://tools.ietf.org/html/draft-ietf-tls-tls13-24

[0] https://www.openssl.org/

In 2018, we examined how the new generation of formal methods can cope with the MAA case study. We specified the MAA in both LOTOS and LNT and checked these specifications using the CADP tools. The C code generated by the CADP compilers was executed w.r.t. a set of reference MAA test vectors, as well as supplementary test vectors devised to improve the coverage of byte permutations and message segmentation. This enabled us to detect and correct several errors in the reference test vectors given in the ISO 8730 and ISO 8731-2 standards. This work led to a publication in an international workshop [22].

6.5.6. Other Case Studies

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

Based on the work described above, the demo examples of the CADP toolbox have been enriched. Two new demo examples have been added: demo_06 (Transport Layer Security v1.3 handshake protocol specified in LNT), and demo_11 (a hardware block implementing a Dynamic Task Dispatcher). The demo_12 (Message Authenticator Algorithm) is now documented in a publication [22]. The demo_17 (distributed leader election protocol) has been converted from LOTOS to LNT. Finally, most existing demo examples have been updated to reflect the evolution of the MCL v3 and SVL languages.

DEDUCTEAM Project-Team

7. New Results

7.1. $\lambda\Pi$-calculus modulo theory

Gilles Dowek, Jean-Pierre Jouannaud and Jiaxiang Liu have started a program for developing new techniques for proving confluence of dependently typed theories, which do not rely on termination. These results have been presented at Types 2016, and will be submitted to a Journal early 2019. Target applications for these techniques are encodings of the Calculus of inductive constructions with polymorphic universes in the $\lambda\Pi$-calculus modulo theory.

Frédéric Blanqui has published in the Journal of Functional Programming a long article synthesizing his work on the use of size annotations for proving termination [12]. 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, 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.

Size-change termination is a technique introduced for first-order functional programs. In [16], Frédéric Blanqui and Guillaume Genestier show how it can be used to study the termination of higher-order rewriting in the $\lambda\Pi$-calculus modulo theory.

Dependency pairs are a key concept at the core of modern automated termination provers for first-order term rewrite systems. In [22], Frédéric Blanqui, Guillaume Genestier and Olivier Hermant introduced an extension of this technique for a large class of dependently-typed higher-order rewrite systems. This improves previous results by Wahlstedt on one hand and Frédéric Blanqui on the other hand to strong normalization and non-orthogonal rewrite systems. This new criterion has been implemented in the type-checker DEDUKTI.

7.2. Dedukti

Frédéric Blanqui and Guillaume Genestier have formally defined the operational semantics of DEDUKTI 2.5, showing some problems with non left-linear rewrite rules.

Rodolphe Lepigre, Frédéric Blanqui and Franck Slama developed a new version of DEDUKTI, available on https://github.com/Deducteam/lambdapi, with meta-variables and a small set of tactics in order to be able to build DEDUKTI proofs interactively.

Aristomenis-Dionysios Papadopoulos has added a rewrite tactic in the style of Ssreflect [27].

Emilio Gallego added an LSP server for communicating with editors.

Ismail Lachheb has developed a plugin for DEDUKTI based on the LSP protocol into the Atom editor [25].

Guillaume Burel added support for polarized Deduction modulo theory in DEDUKTI.

Quentin Ye has developed an algorithm to compare λ-terms. The main point was to take sharing into account, so as to relate the complexity with the space used to represent the term, rather than with the size of the term. He has implemented this algorithm in the DEDUKTI codebase. He has also run his algorithm on examples that show an exponential speed-up compared to the naive algorithm [21].

7.3. Theories

Gaspard Férey and François Thiré defined a new encoding for Cumulative type systems (CTS) in the $\lambda\Pi$-calculus modulo theory, extending the work of Ali Assaf's PhD [28]. This encoding relies on explicit subtyping which requires additional computational rules. It provides a way to encode a larger class of CTS, which sheds a new light on the computational content of explicit subtyping. This encoding should be extendable to express more advanced features such as universe polymorphism in the Calculus of Inductive Construction, a first step to have a faithful encoding of the COQ system. The encoding has been proven correct under the hypothesis that the computational rules are confluent.

François Thiré redesigned the tool UNIVERSO, so that it can be used for a larger class of CTS. The specification for UNIVERSO can be given by rewrite rules which makes UNIVERSO much easier to use. This tool is a first step to have an automatic chain of translations to translate proofs in the encoding of MATITA to STT$\forall_{\beta\delta}$.which would make these proofs interoperable with 5 different systems.

François Thiré changed the encoding provided by KRAJONO to integrate some ideas of the encoding discussed above. This encoding is compatible with the tool UNIVERSO.

Gaspard Férey updated the COQINE software to translate COQ's 8.8 version. In this version, the standard library relies on universe polymorphism so partial support for the translation of this feature was integrated. Since encodings of the many features of Coq (inductive constructions, floating universes, several kinds of universe polymorphisms, etc) are a current work in progress, the software was made parameterizable to allow experimentations of multiple encodings of these features.

Gaspard Férey showcased an encoding of the Calculus of Inductive Constructions (CiC) relying on associative-commutative (AC) rewriting on the arithmetic library translated from MATITA. This practical experiment shows the limitations of AC-rewriting (as implemented in DEDUKTI) in terms of performance and the need for special care when defining encodings relying on this feature.

Guillaume Burel began to write a tool translating SAT proof traces in LRAT format into DEDUKTI proofs. The main issue was that steps in LRAT traces are not logical consequences of previous clauses but only preserve provability.

Mohamed Yacine El Haddad developed a tool to extract TPTP problems from a TSTP trace (generated by automated theorem provers) and reconstruct the proof of the trace in DEDUKTI format.

Bruno Barras has started to develop a model of Homotopy Type Theory (HoTT) in DEDUKTI. This is basically a presheaf model, where the choice of the base category leads either to the simplicial sets model or to the cubical model of HoTT. This construction generalizes the setoid model construction [2] to an arbitrary dimension. Since this involves encoding notions of category theory, the rewriting feature of DEDUKTI is intensively used to represent, among others, the associativity of morphism composition, or the naturality conditions.

Guillaume Bury has proposed an automation-friendly set theory for the B method. This theory is expressed using first order logic extended to polymorphic types and rewriting. Rewriting is introduced along the lines of deduction modulo theory, where axioms are turned into rewrite rules over both propositions and terms. This work has been published in [30].

7.4. Interoperability

François Thiré has defined in DEDUKTI a constructive version of simple type theory with prenex polymorphism: STT$\forall_{\beta\delta}$. This work has been published at the LFMTP workshop in [15]. STT$\forall_{\beta\delta}$ has been used to encode an arithmetic library able to prove little Fermat's theorem. Then these proofs has been exported to different systems that are: COQ, MATITA, LEAN and OPENTHEORY. Gilles Dowek, César Muñoz, and François Thiré have developed a translation of STT$\forall_{\beta\delta}$ to PVS.

Then, Walid Moustaoui and François Thiré have built a website called LOGIPEDIA which allows the user to inspect this arithmetic library and the user can download the proof of this theorem to one of the systems mentioned above.

7.5. Drags

Shared and cyclic structures are very common in both programming and proving, which requires generalizing term rewriting techniques to graphs. Jean-Pierre Jouannaud and Nachum Dershowitz have introduced a very general class of multigraphs, called drags, equipped with a composition operator $\otimes$ which provides with a rich categorical structure. Rewriting a drag D can then be defined in a very simple way, by writing D as the composition of a left-hand side of rules L and a context C, and then replacing L by R, the right-hand side of the rule, which yields the rewritten drag $R \otimes C$. The fundamental aspects of the algebra of drags have been presented at TERMGRAPH'2018 and have also been submitted to a special issue of TCS. Termination of drag rewriting in investigated in [20].

7.6. SCTL

Gilles Dowek, Liu Jian, and Ying Jiang have reworked the presentation of CTL in sequent calculus proposed by Gilles Dowek and Ying Jiang in 2012 and provided an implementation of it. This work has been published in [13].

GALLINETTE Project-Team

6. New Results

6.1. Logical Foundations of Programming Languages

Participants: Rémi Douence, Ambroise Lafont, Étienne Miquey, Xavier Montillet, Guillaume Munch-Maccagnoni, Nicolas Tabareau, Pierre Vial.

6.1.1. Classical Logic

6.1.1.1. A sequent calculus with dependent types for classical arithmetic.

In a recent paper, Herbelin developed a calculus $dPA\omega$ in which constructive proofs for the axioms of countable and dependent choices could be derived via the encoding of a proof of countable universal quantification as a stream of it components. However, the property of normalisation (and therefore the one of soundness) was only conjectured. The difficulty for the proof of normalisation is due to the simultaneous presence of dependent dependent types (for the constructive part of the choice), of control operators (for classical logic), of coinductive objects (to encode functions of type $N \to A$ into streams $(a_0, a_1, \cdots)$) and of lazy evaluation with sharing (for these coinductive objects).Building on previous works, we introduce in [14], [26] a variant of $dPA\omega$ presented as a sequent calculus. On the one hand, we take advantage of a variant of Krivine classical realisability we developed to prove the normalisation of classical call-by-need. On the other hand, we benefit of dL, a classical sequent calculus with dependent types in which type safety is ensured using delimited continuations together with a syntactic restriction. By combining the techniques developed in these papers, we manage to define a realisability interpretation à la Krivine of our calculus that allows us to prove normalisation and soundness.

6.1.1.2. Realisability Interpretation and Normalisation of Typed Call-by-Need λ-calculus With Control.

In [13], we define a variant of realisability where realisers are pairs of a term and a substitution. This variant allows us to prove the normalisation of a simply-typed call-by-need λ-calculus with control due to Ariola et al. Indeed, in such call-by-need calculus, substitutions have to be delayed until knowing if an argument is really needed. In a second step, we extend the proof to a call-by-need λ-calculus equipped with a type system equivalent to classical second-order predicate logic, representing one step towards proving the normalisation of the call-by-need classical second-order arithmetic introduced by the second author to provide a proof-as-program interpretation of the axiom of dependent choice.

6.1.2. Lambda Calculus

6.1.2.1. Every λ-Term is Meaningful for the Infinitary Relational Model.

Infinite types and formulas are known to have really curious and unsound behaviours. For instance, they allow to type Ω, the auto-autoapplication and they thus do not ensure any form of normalisation/productivity. Moreover, in most infinitary frameworks, it is not difficult to define a type R that can be assigned to every λ-term. However, these observations do not say much about what coinductive (i.e. infinitary) type grammars are able to provide: it is for instance very difficult to know what types (besides R) can be assigned to a given term in this setting. In [17], we begin with a discussion on the expressivity of different forms of infinite types. Then, using the resource-awareness of sequential intersection types (system S) and tracking, we prove that infinite types are able to characterise the arity of every λ-terms and that, in the infinitary extension of the relational model, every term has a " meaning " i.e. a non-empty denotation. From the technical point of view, we must deal with the total lack of guarantee of productivity for typable terms: we do so by importing methods inspired by first order model theory.

6.1.2.2. High-level signatures and initial semantics.

In [9], we present a device for specifying and reasoning about syntax for datatypes, programming languages, and logic calculi. More precisely, we consider a general notion of signature for specifying syntactic constructions. Our signatures subsume classical algebraic signatures (i.e., signatures for languages with variable binding, such as the pure lambda calculus) and extend to much more general examples. In the spirit of Initial Semantics, we define the syntax generated by a signature to be the initial object—if it exists—in a suitable category of models. Our notions of signature and syntax are suited for compositionality and provide, beyond the desired algebra of terms, a well-behaved substitution and the associated inductive/recursive principles. Our signatures are general in the sense that the existence of syntax is not automatically guaranteed. In this work, we identify a large class of signatures which do generate a syntax. This paper builds upon ideas from a previous attempt by Hirschowitz-Maggesi (FICS 2012), which, in turn, was directly inspired by some earlier work of Ghani-Uustalu and Matthes-Uustalu. The main results presented in the paper are computer-checked within the UniMath system.

6.1.3. Models of programming languages mixing effects and resources

6.1.3.1. A resource modality for RAII

Systems programming languages C++11 and Rust have developed techniques and idioms for the safe management of resources called *"Resource acquisition is initialisation"* (RAII) and *move semantics*. We have related resources from systems programming to the notion of resource put forward by linear logic, by giving a construction in terms of categorical semantics for a resource modality that model RAII and move semantics. This work was presented by at the workshop LOLA 2018 in Oxford [20].

6.1.3.2. Resource polymorphism

Thanks to a new logical and semantic understanding of resource-management techniques in systems programming languages, we have proposed [27] a design for an extension of functional programming language towards systems programming, centred on the OCaml language, and based on a notion of resource polymorphism inspired by the C++11 language and by the works on polarisation in proof theory.

6.1.4. Distributed Programming

6.1.4.1. Chemical foundations of distributed aspects.

Distributed applications are challenging to program because they have to deal with a plethora of concerns, including synchronisation, locality, replication, security and fault tolerance. Aspect-oriented programming (AOP) is a paradigm that promotes better modularity by providing means to encapsulate cross-cutting concerns in entities called aspects. Over the last years, a number of distributed aspect-oriented programming languages and systems have been proposed, illustrating the benefits of AOP in a distributed setting. Chemical calculi are particularly well-suited to formally specify the behaviour of concurrent and distributed systems. The join calculus is a functional name-passing calculus, with both distributed and object-oriented extensions. It is used as the basis of concurrency and distribution features in several mainstream languages like C# (Polyphonic C#, now Cω), OCaml (JoCaml), and Scala Joins. Unsurprisingly, practical programming in the join calculus also suffers from modularity issues when dealing with crosscutting concerns. We propose the Aspect Join Calculus [8], an aspect-oriented and distributed variant of the join calculus that addresses crosscutting and provides a formal foundation for distributed AOP. We develop a minimal aspect join calculus that allows aspects to advise chemical reactions. We show how to deal with causal relations in pointcuts and how to support advanced customisable aspect weaving semantics.

6.2. Type Theory and Proof Assistants

Participants: Simon Boulier, Eric Finster, Gaëtan Gilbert, Pierre-Marie Pédrot, Nicolas Tabareau, Théo Winterhalter.

6.2.1. Type Theory

6.2.1.1. Effects in Type Theory.

In [16] , we define the exceptional translation, a syntactic translation of the Calculus of Inductive Constructions (CIC) into itself, that covers full dependent elimination. The new resulting type theory features call-by-name exceptions with decidable type-checking and canonicity, but at the price of inconsistency. Then, noticing parametricity amounts to Kreisel's realisability in this setting, we provide an additional layer on top of the exceptional translation in order to tame exceptions and ensure that all exceptions used locally are caught, leading to the parametric exceptional translation which fully preserves consistency. This way, we can consistently extend the logical expressivity of CIC with independence of premises, Markov's rule, and the negation of function extensionality while retaining η-expansion. As a byproduct, we also show that Markov's principle is not provable in CIC. Both translations have been implemented in a Coq plugin, which we use to formalise the examples.

6.2.1.2. Eliminating Reflection from Type Theory.

Type theories with equality reflection, such as extensional type theory (ETT), are convenient theories in which to formalise mathematics, as they make it possible to consider provably equal terms as convertible. Although type-checking is undecidable in this context, variants of ETT have been implemented, for example in NuPRL and more recently in Andromeda. The actual objects that can be checked are not proof-terms, but derivations of proof-terms. This suggests that any derivation of ETT can be translated into a typecheckable proof term of intensional type theory (ITT). However, this result, investigated categorically by Hofmann in 1995, and 10 years later more syntactically by Oury, has never given rise to an effective translation. In [18], we provide the first syntactical translation from ETT to ITT with uniqueness of identity proofs and functional extensionality. This translation has been defined and proven correct in Coq and yields an executable plugin that translates a derivation in ETT into an actual Coq typing judgment. Additionally, we show how this result is extended in the context of homotopy to a two-level type theory.

6.2.1.3. Foundations of Dependent Interoperability.

Full-spectrum dependent types promise to enable the development of correct-by-construction software. However, even certified software needs to interact with simply-typed or untyped programs, be it to perform system calls, or to use legacy libraries. Trading static guarantees for runtime checks, the dependent interoperability framework provides a mechanism by which simply-typed values can safely be coerced to dependent types and, conversely, dependently-typed programs can defensively be exported to a simply-typed application. In [2], we give a semantic account of dependent interoperability. Our presentation relies on and is guided by a pervading notion of type equivalence, whose importance has been emphasised in recent work on homotopy type theory. Specifically, we develop the notion of type-theoretic partial Galois connections as a key foundation for dependent interoperability, which accounts for the partiality of the coercions between types. We explore the applicability of both monotone and antitone type-theoretic Galois connections in the setting of dependent interoperability. A monotone partial Galois connection enforces a translation of dependent types to runtime checks that are both sound and complete with respect to the invariants encoded by dependent types. Conversely, picking an antitone partial Galois connection instead lets us induce weaker, sound conditions that can amount to more efficient runtime checks. Our framework is developed in Coq; it is thus constructive and verified in the strictest sense of the terms. Using our library, users can specify domain-specific partial connections between data structures. Our library then takes care of the (sometimes, heavy) lifting that leads to interoperable programs. It thus becomes possible, as we shall illustrate, to internalise and hand-tune the extraction of dependently-typed programs to interoperable OCaml programs within Coq itself.

6.2.1.4. Equivalences for Free: Univalent Parametricity for Effective Transport.

Homotopy Type Theory promises a unification of the concepts of equality and equivalence in Type Theory, through the introduction of the univalence principle. However, existing proof assistants based on type theory treat this principle as an axiom, and it is not yet clear how to extend them to handle univalence internally. In [7], we propose a construction grounded on a univalent version of parametricity to bring the benefits of univalence to the programmer and prover, that can be used on top of existing type theories. In particular, univalent parametricity strengthens parametricity to ensure preservation of type equivalences. We present a lightweight framework implemented in the Coq proof assistant that allows the user to transparently transfer definitions and theorems for a type to an equivalent one, as if they were equal. Our approach handles both type and term dependency. We study how to maximise the effectiveness of these transports in terms of computational behaviour, and identify a fragment useful for certified programming on which univalent transport is guaranteed to be effective. This work paves the way to easier-to-use environments for certified programming by supporting seamless programming and proving modulo equivalences.

6.2.1.5. Special Issue on Homotopy Type Theory and Univalent Foundations.

The preface [4] introduces the first special issue out of a series of workshops on Homotopy Type Theory and Univalent Foundations. This recent area of research finds its roots in the seminal work of Martin Hofmann and Thomas Streicher on the structure of Martin-Löf identity types. But the main research program has been foreseen by Vladimir Voevodsky, who, from its initial motivation of formalising his results in homotopy theory, has initiated what is now called the univalent foundations program. Borrowing ideas from homotopy theory, the goal of the univalent foundations program is to leverage dependent Type Theory to a formal framework that could replace Set Theory for the foundations of mathematics. This special issue gathers research contributions of some of the most prominent researchers of the field.

6.2.1.6. Goodwillie's Calculus of Functors and Higher Topos Theory

In [1], we develop an approach to Goodwillie's calculus of functors using the techniques of higher topos theory. Central to our method is the introduction of the notion of fiberwise orthogonality, a strengthening of ordinary orthogonality which allows us to give a number of useful characterisations of the class of n-excisive maps. We use these results to show that the pushout product of a P_n-equivalence with a P_m-equivalence is a P_{m+n+1}-equivalence. Then, building on our previous work, we prove a Blakers-Massey type theorem for the Goodwillie tower. We show how to use the resulting techniques to rederive some foundational theorems in the subject, such as delooping of homogeneous functors.

6.2.2. Proof Assistants

6.2.2.1. Typed Template Coq – Certified Meta-Programming in Coq.

Template-Coq [19], [10] is a plugin for Coq, originally implemented by Malecha, which provides a reifier for Coq terms and global declarations , as represented in the Coq kernel, as well as a denotation command. Initially, it was developed for the purpose of writing functions on Coq's AST in Gallina. Recently, it was used in the CertiCoq certified compiler project, as its front-end language, to derive parametricity properties, and to extract Coq terms to a CBV λ-calculus. However, the syntax lacked semantics, be it typing semantics or operational semantics, which should reflect, as formal specifications in Coq, the semantics of Coq's type theory itself. The tool was also rather bare bones, providing only rudimentary quoting and unquoting commands. We generalise it to handle the entire Calculus of Inductive Constructions (CIC), as implemented by Coq, including the kernel's declaration structures for definitions and inductives, and implement a monad for general manipulation of Coq's logical environment. We demonstrate how this setup allows Coq users to define many kinds of general purpose plugins, whose correctness can be readily proved in the system itself, and that can be run efficiently after extraction. We give a few examples of implemented plugins, including a parametricity translation. We also advocate the use of Template-Coq as a foundation for higher-level tools.

6.2.2.2. Definitional Proof-Irrelevance without K.

Definitional equality—or conversion—for a type theory with a decidable type checking is the simplest tool to prove that two objects are the same, letting the system decide just using computation. Therefore, the more things are equal by conversion, the simpler it is to use a language based on type theory. Proof-irrelevance, stating that any two proofs of the same proposition are equal, is a possible way to extend conversion to make a type theory more powerful. However, this new power comes at a price if we integrate it naively, either by making type checking undecidable or by realising new axioms—such as uniqueness of identity proofs (UIP)—that are incompatible with other extensions, such as univalence. In [3], taking inspiration from homotopy type theory, we propose a general way to extend a type theory with definitional proof irrelevance, in a way that keeps type checking decidable and is compatible with univalence. We provide a new criterion to decide whether a proposition can be eliminated over a type (correcting and improving the so-called singleton elimination of Coq) by using techniques coming from recent development on dependent pattern matching without UIP. We show the generality of our approach by providing implementations for both Coq and Agda, both of which are planned to be integrated in future versions of those proof assistants.

6.3. Program Certifications and Formalisation of Mathematics

Participants: Danil Annenkov, Assia Mahboubi, Étienne Miquey.

6.3.1. Certified Compilation of Financial Contracts.

In [11], we present an extension to a certified financial contract management system that allows for templated declarative financial contracts and for integration with financial stochastic models through verified compilation into so-called payoff-expressions. Such expressions readily allow for determining the value of a contract in a given evaluation context, such as contexts created for stochastic simulations. The templating mechanism is useful both at the contract specification level, for writing generic reusable contracts, and for reuse of code that, without the templating mechanism, needs to be recompiled for different evaluation contexts. We report on the effect of using the certified system in the context of a GPGPU-based Monte Carlo simulation engine for pricing various over-the-counter (OTC) financial contracts. The full contract-management system, including the payoff-language compilation, is verified in the Coq proof assistant and certified Haskell code is extracted from our Coq development along with Futhark code for use in a data-parallel pricing engine.

6.3.2. Static interpretation of higher-order modules in Futhark: functional GPU programming in the large.

In [12], we present a higher-order module system for the purely functional data-parallel array language Futhark. The module language has the property that it is completely eliminated at compile time, yet it serves as a powerful tool for organising libraries and complete programs. The presentation includes a static and a dynamic semantics for the language in terms of, respectively, a static type system and a provably terminating elaboration of terms into terms of an underlying target language. The development is formalised in Coq using a novel encoding of semantic objects based on products, sets, and finite maps. The module language features a unified treatment of module type abstraction and core language polymorphism and is rich enough for expressing practical forms of module composition.

6.3.3. Formalising Implicative Algebras in Coq.

In [15], we present a Coq formalisation of Alexandre Miquel's implicative algebras, which aim at providing a general algebraic framework for the study of classical realisability models. We first give a self-contained presentation of the underlying implicative structures, which roughly consists of a complete lattice equipped with a binary law representing the implication. We then explain how these structures can beturned into models by adding separators, giving rise to the so-called implicative algebras. Additionally, we show how they generalise Boolean and Heyting algebras as well as the usual algebraic structures used in the analysis of classical realisability.

6.3.4. Formally Verified Approximations of Definite Integrals.

Finding an elementary form for an antiderivative is often a difficult task, so numerical integration has become a common tool when it comes to making sense of a definite integral. Some of the numerical integration methods can even be made rigorous: not only do they compute an approximation of the integral value but they also bound its inaccuracy. Yet numerical integration is still missing from the toolbox when performing formal proofs in analysis. In [5], we present an efficient method for automatically computing and proving bounds on some definite integrals inside the Coq formal system. Our approach is not based on traditional quadrature methods such as Newton-Cotes formulas. Instead, it relies on computing and evaluating antiderivatives of rigorous polynomial approximations, combined with an adaptive domain splitting. Our approach also handles improper integrals, provided that a factor of the integrand belongs to a catalog of identified integrable functions. This work has been integrated to the CoqInterval library.

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].

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:

- A new built-in function, `__builtin_ais_annot` makes it easy to transfer annotations (also known as flow facts) written at the source code level in AbsInt's aiS annotation language all the way down to the level of the generated machine code. The aiT static analyzer for Worst-Case Execution Times, which operates at the machine code level, can then take advantage of these annotations to produce better WCET estimates.
- In preparation for a qualification with respect to industry standards for avionics software, conformance with the ISO C 1999 and ISO C 2011 standards was improved, with the addition of many diagnostics required by the standards.
- Performance of the generated code was slightly improved via changes to the heuristics for function inlining and for instruction selection.
- The semantic modeling of external function calls was made more precise, reflecting the fact that these functions can destroy some registers and some stack locations.

We released three versions of CompCert incorporating these improvements: version 3.2 in January 2018, version 3.3 in May 2018, and version 3.4 in September 2018.

Two papers on CompCert were presented at conferences. The first paper, with Daniel Kästner as lead author, was presented at the 2018 ERTS congress [22]. 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. The second paper, with Bernhard Schommer as lead author, was presented at the 2018 WCET workshop [23]. It describes the `__builtin_ais_annot` source-level annotation mechanism mentioned above and its uses to help WCET analysis.

7.1.2. Verified code generation in the polyhedral model

Participants: Nathanaël Courant, Xavier Leroy.

The polyhedral model is a high-level intermediate representation for loop nests iterating over arrays and matrices, as found in numerical code. It supports a great many loop optimizations (fusion, splitting, interchange, blocking, etc) in a uniform, mathematically-elegant manner.

Nathanaël Courant, as part of his MPRI Master's internship and under Xavier Leroy's supervision, developed a Coq formalization of the polyhedral model. He then implemented and proved correct in Coq a code generator that produces efficient sequential code from an optimized polyhedral representation. Code generation is a delicate part of polyhedral compilation, involving complex, error-prone algorithms. Nathanaël Courant's verified code generator includes the major algorithms from Cédric Bastoul's reference paper [31]. The Coq specifications and proofs are available at https://github.com/Ekdohibs/PolyGen.

7.1.3. *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 continued work on 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 using a custom binary analysis tool. This tool finds missed optimizations by comparing 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.

The system has found previously unreported missing optimizations in the GCC, Clang, and CompCert compilers. An article [19] was presented at the 27th International Conference on Compiler Construction (CC 2018), where it was honored with the Best Paper Award.

7.1.4. *A verified model of register aliasing in CompCert*

Participants: Gergö Barany, Xavier Leroy.

Some CPU architectures such as ARM feature register aliasing: Each of its 64-bit floating-point registers can also be accessed as two separate 32-bit halves. Modifying a superregister changes (invalidates) the data stored in subregisters and vice versa, but this behavior was not yet modeled in CompCert's semantics.

We continued work on re-engineering much of CompCert's 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 was presented at the workshop on Syntax and Semantics of Low-Level Languages (LOLA 2018) [25].

7.2. Language design and type systems

7.2.1. *Refactoring with ornaments in ML*

Participants: Thomas Williams, Lucas Baudin, Didier Rémy.

Thomas Williams, Lucas Baudin, and Didier Rémy have been working on refactoring and other transformations of ML programs based on mixed ornamentation and disornamentation. Ornaments have been introduced as a way of describing changes in data type definitions that can reorganize or add 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.

Williams and Rémy improved the formalisation of the lifting framework: 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 to 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. This work has been presented at POPL 2018 [12]. More technical details appear in a research report [43].

Lucas Baudin and Dider Rémy also studied the inverse transformation, disornamentation, which allows removing pieces of information from a data structure and adjusting the code accordingly. They showed that the framework of ornamentation can also be used to allow mixed ornamentation and disornamentation transformations. They also designed a new patch language to describe in a more robust manner how the code must be modified during such transformations. This enables a new class of applications, such as maintaining two views of a data structure in sync. For example, the location information in an abstract syntax tree, which is used to report error messages but obfuscates the code, can be projected away, leading to a simpler version of the code, which can then be modified and often automatically reornamented into the richer version of the code with locations. Disonamentation has been presented by Lucas Baudin at the ML 2018 workshop. Ornamentation, including mixed disornamentation, has also been presented at the MSFP 2018 workshop in Oxford.

A small prototype with ornamentation has been written by Thomas Williams and extended with disornamentation by Lucas Baudin. Thomas Williams has also started developing a new version of the prototype that will handle most of the OCaml language.

7.3. Shared-memory concurrency

7.3.1. The Linux Kernel Memory Model

Participants: Luc Maranget, Jade Alglave [University College London & ARM Ltd], 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 optimisations 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 a publication on languages in an international conference. A multi-year effort to define a weak memory model for the Linux Kernel has yielded a scholarly paper [18] presented at the *Architectural Support for Programming Languages and Operating Systems* (ASPLOS) conference in March 2018. The article describes a formal model, the *Linux Kernel Memory Model* (LKMM), which defines how Linux kernel programs are supposed to behave. The model, a CAT model, can be simulated using the **herd** simulator, allowing programmers to experiment and develop intuitions. The model was tested against hardware and refined in consultation with Linux maintainers. Finally, the ASPLOS paper formalizes the *fundamental law of the Read-Copy-Update synchronization mechanism* and proves that one of its implementations satisfies this law. It is worth noting that the LKMM is now part of the Linux kernel source (in the `tools/`) section). Luc Maranget and his co-authors are the official maintainers of this document.

7.3.2. The ARMv8 and RISC-V memory model

Participants: Will Deacon [ARM Ltd], Luc Maranget, Jade Alglave [University College London & ARM Ltd].

Jade Alglave and Luc Maranget are working on a mixed-size version of the ARMv8 memory model. This model builds on the `aarch64.cat` model authored last year by Will Deacon (ARM Ltd). This ongoing work is subject to IP restrictions which we hope to lift next year.

Luc Maranget is an individual member of the memory model group of the RISC-V consortium (https://riscv.org/). Version V2.3 of the User-Level ISA Specification is now complete and should be released soon. This version features the first occurrence of a detailed memory model expressed in English, as well as its transliteration in CAT authored by Luc Maranget.

7.3.3. Work on diy

Participant: Luc Maranget.

This year, new synchronisation primitives were added to the Linux kernel memory model; ARMv8 atomic instructions were added; and more.

A more significant improvement is the introduction of *mixed-size* accesses. The tools can now handle a new view of memory, where memory is made up of elementary cells (typically *bytes*) that can be read or written as groups of contiguous cells (typically up to *quadwords* of 8 bytes). This preliminary work paves the way to the simulation of more elaborate memory models.

7.3.4. Unifying axiomatic and operational weak memory models

Participants: Jean-Marie Madiot, Jade Alglave [University College London & ARM Ltd], Simon Castellan [Imperial College London].

Modern multi-processors optimize the running speed of programs using a variety of techniques, including caching, instruction reordering, and branch speculation. While those techniques are perfectly invisible to sequential programs, such is not the case for concurrent programs that execute several threads and share memory: threads do not share at every point in time a single consistent view of memory. A *weak memory model* offers only weak consistency guarantees when reasoning about the permitted behaviors of a program. Until now, there have been two kinds of such models, based on different mathematical foundations: axiomatic models and operational models.

Axiomatic models explicitly represent the dependencies between the program and memory actions. These models are convenient for causal reasoning about programs. They are also well-suited to the simulation and testing of *hardware* microprocessors.

Operational models represent program states directly, thus can be used to reason on programs: program logics become applicable, and the reasoning behind nondeterministic behavior is much clearer. This makes them preferable for reasoning about *software*.

Jean-Marie Madiot has been collaborating with weak memory model expert Jade Alglave and concurrent game semantics researcher Simon Castellan in order to unify these styles, in a way that attempts to combine the best of both approaches. The first results are a formalisation of TSO-style architectures using partial-order techniques similar to the ones used in game semantics, and a proof of a stronger-than-state-of-art “data-race freedom” theorem: well-synchronised programs can assume a strong memory model. These results have been submitted for publication.

This is a first step towards tractable verification of concurrent programs, combining software verification using concurrent program logics, in the top layer, and hardware testing using weak memory models, in the bottom layer. Our hope is to leave no unverified gap between software and hardware, even (and especially) in the presence of concurrency.

7.3.5. Granularity control for parallel programs

Participants: Umut Acar, Vitaly Aksenov, Arthur Charguéraud, Adrien Guatto [Université Paris Diderot], Mike Rainey, Filip Sieczkowski [University of Wrocław].

This year, the DeepSea team continued their work on granularity control techniques for parallel programs.

A first line of research is based on the use of programmer-supplied asymptotic complexity functions, combined with runtime measurements. This work first appeared at PPoPP 2018 [16] in the form of a brief announcement, and was subsequently accepted for publication at PPoPP 2019 as a full paper.

A second line of research, known as *heartbeat scheduling*, is based on instrumenting the runtime system so that parallel function calls are initially executed as normal function calls, by pushing a frame on the stack, and subsequently can be promoted and become independent threads. This research has been presented at PLDI 2018 [14].

7.3.6. Theory and analysis of concurrent algorithms

Participant: Vitaly Aksenov.

Vitaly Aksenov, in collaboration with Petr Kuznetsov (Télécom ParisTech) and Anatoly Shalyto (ITMO University), proved that no wait-free linearizable implementation of a stack using read, write, compare & swap and fetch & add operations can be help-free. This proof corrects a mistake in an earlier proof by Censor-Hillel et al. The result was published at the the International Conference on Networked Systems (NETYS 2018) [17].

Vitaly Aksenov, in collaboration with Dan Alistarh (IST Austria) and Petr Kuznetsov (Télécom ParisTech), worked on performance prediction for coarse-grained locking. They describe a simple model that can be used to predict the throughput of coarse-grained lock-based algorithms. They show that their model works well for CLH locks, and thus can be expected to work for other popular lock designs such as TTAS or MCS. This work appeared as a brief announcement at PODC 2018 [16].

The aforementioned results by Vitaly Aksenov are also covered in his Ph.D. manuscript [11].

7.4. The OCaml language and system

7.4.1. The OCaml system

Participants: Damien Doligez, Armaël Guéneau, Xavier Leroy, Luc Maranget, David Allsop [University of Cambridge], Florian Angeletti, Frédéric Bour [Facebook], Stephen Dolan [University of Cambridge], Alain Frisch [Lexifi], Jacques Garrigue [University of Nagoya], Sébastien Hinderer, Nicolás Ojeda Bär [Lexifi], Thomas Refis [Jane Street], Gabriel Scherer [team Parsifal], Mark Shinwell [Jane Street], Leo White [Jane Street], Jeremy Yallop [University of Cambridge].

This year, we released three versions of the OCaml system: versions 4.06.1 and 4.07.1 are minor releases that fix 7 and 8 issues, respectively; version 4.07.0 is a major release that introduces many improvements in usability and performance, and fixes about 40 issues. The main novelties are:

- The standard library modules were reorganized to appear as sub-modules of a new `Stdlib` module. The purpose of this reorganization is to facilitate the addition of new standard library modules while minimize risks of conflicts with user modules of the same name.
- Modules `Float` (floating-point operations) and `Seq` (sequences) were added to the standard library, taking advantage of the new organization mentioned above.
- Since 4.01, it has been possible to select a variant constructor or record field from a sub-module that is not opened in the current scope, if type information is available at the point of use. This now also works for GADT constructors.
- The GC now handles the accumulation of custom blocks in the minor heap better. This solves some memory-usage issues observed in code which allocates a large amount of small custom blocks, typically small bigarrays.

7.4.2. Package management infrastructure

Participant: Damien Doligez.

This year, Damien Doligez has worked on the `opamcheck` tool, which is designed to check the compatibility of different versions of OCaml on the whole code base of `opam`, OCaml's package manager. As a by-product of this work, he has proposed numerous fixes to the `opam` package repository and to its dependency graph.

7.4.3. Work on the compiler's test suite and build system

Participant: Sébastien Hinderer.

In 2018, Sébastien Hinderer has worked on the OCaml compiler's test suite. More precisely, he has finished porting over 800 tests in the compiler's test suite so that they can be run by the tool `ocamltest`, developed by Sébastien earlier. To achieve this, it has been necessary to extend both `ocamltest` and the domain-specific language that is used to describe how tests should be executed.

In addition, Sébastien has fixed and properly documented the procedure that is used to bootstrap the OCaml compiler. Being able to compile the compiler using itself is an important feature: it is crucial, for instance, when the compiler is released. In addition to fixing the bootstrap procedure, Sébastien has introduced a way to test this procedure through continuous integration, which guarantees that it will not be broken again in the future.

Finally, Sébastien has continued to improve and refactor the compiler's build system, and, most importantly, has replaced the hand-written configuration script by an `autoconf`-generated one, which will be part of the upcoming 4.08 release of OCaml. This represents an important step towards the ability to produce cross-compilers for OCaml, which has been a long-standing issue for the whole OCaml community.

7.4.4. Optimizing OCaml for satisfiability problems

Participants: Sylvain Conchon [LRI, Univ. Paris-Saclay], Albin Coquereau [ENSTA-ParisTech], Mohamed Iguernlala [OCamlPro], Fabrice Le fessant [OCamlPro], Michel Mauny.

This work aims at improving the performance of the Alt-Ergo SMT solver, which is implemented in OCaml. For safety reasons, and to ease reasoning about its algorithms, the implementation of Alt-Ergo uses 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.

Following our previous work on optimizing Alt-Ergo's built-in SAT solver, some efforts were needed to enable the comparison of our solver with other SMT solvers. We developed an OCaml library for parsing and type-checking SMT-LIB2. Since Alt-Ergo natively uses a polymorphic typing discipline, and since the community needs such advanced features, we proposed an extension of the SMT-LIB2 syntax where functions may be polymorphic.

The resulting new version of Alt-Ergo was presented at the 2018 SMT Workshop in Oxford [33]. Comparisons of Alt-Ergo with other SMT solvers, mainly developed in C++, took place during the competition that is associated with the workshop. They showed that Alt-Ergo's performance is similar to that of its competitors.

Albin Coquereau's Ph.D. defense is planned for Spring 2019.

7.4.5. Improvements in Menhir

Participant: François Pottier.

In 2018, the OCaml parser of the OCaml compiler was migrated from `ocamlyacc` to Menhir, at last. François Pottier took this opportunity to partially clean up the parser, reducing redundancy by taking advantage of Menhir's features. In the future, we hope to continue to work on the OCaml parser by improving the quality of its syntax error messages.

This cleanup work was also an occasion to revisit Menhir's grammar description language: François Pottier designed and implemented a new input syntax for Menhir, which seems slightly more powerful and elegant than the previous syntax.

7.5. Software specification and verification

7.5.1. Formal reasoning about asymptotic complexity

Participants: Armaël Guéneau, Arthur Charguéraud [team Camus], François Pottier.

For a couple years, Armaël Guéneau, Arthur Charguéraud, 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. In particular, Armaël has developed in Coq a theory and a set of tactics that allow working with asymptotic complexity bounds. He has presented the main aspects of this work at the conference ESOP 2018 [21]. Furthermore, a key part of the machinery for working with asymptotic complexity bounds has been released as a standalone, reusable Coq library, procrastination. Armaël presented this library at the Coq Workshop in July 2018 [29].

In 2018, Armaël has worked on a more ambitious case study, namely a recent incremental cycle detection algorithm, whose amortized complexity analysis is nontrivial. A machine-checked proof has been completed; a paper is in preparation.

7.5.2. *Time Credits and Time Receipts in Iris*

Participants: Glen Mével, Jacques-Henri Jourdan [CNRS], François Pottier.

From March to August 2018, Glen Mével did an M2 internship at Gallium, where he was co-advised by Jacques-Henri Jourdan (CNRS) and François Pottier. Glen extended the program logic Iris with time credits and time receipts.

Time credits are a well-understood concept, and have been used in several papers already by Armaël Guéneau, Arthur Charguéraud, and François Pottier. However, because Iris is implemented and proved sound inside Coq, extending Iris with time credits requires a nontrivial proof, which Glen carried out, based on a program transformation which inserts "tick" instructions into the code. As an application of time credits, Glen verified inside Iris the correctness of Okasaki's notion of "debits", which allows reasoning about the time complexity of programs that use thunks.

Time receipts are a new concept, which (we showed) allows proving that certain undesirable events, such as integer overflows, cannot occur until a very long time has elapsed. Glen extended Iris with time receipts and proved the soundness of this extension. As an application of time credits and receipts together, Jacques-Henri Jourdan updated Charguéraud and Pottier's earlier verification of the Union-Find data structure [3] and proved that integer ranks cannot realistically overflow, even if they are stored using only $\log W$ bits, where W is the number of bits in a machine word.

This work has been first submitted to POPL 2019, then (after significant revision) re-submitted to ESOP 2019.

7.5.3. *Verified Interval Maps*

Participant: François Pottier.

In the setting of ANR project Vocal, which aims to build a library of verified data structures for OCaml, François Pottier carried out a formal reconstruction of "interval maps". An interval map, a data structure proposed by Bonichon and Cuoq in 2010, represents a set of possible heaps, that is, a set of mappings of integer addresses to abstract values. Interval maps are used in the Frama-C program analysis tool. François Pottier re-implemented this data structure in Coq and carried out a formal verification of its main operations. This work, which represents about 4 months of work, remains unpublished at this time. It would be desirable to publish it and to envision its integration in Frama-C; this however requires further effort.

7.5.4. *Chunked Sequences*

Participants: Émilie Guermeur, Arthur Charguéraud, François Pottier.

In June and July 2018, Émilie Guermeur, an undergraduate student at Carnegie Mellon University (Pittsburgh, USA) did a 6-week internship, co-advised by Arthur Charguéraud and François Pottier. She wrote a full-fledged OCaml implementation of "chunked sequences", a data structure which offers an efficient representation of sequences of elements. This data structure exists in two forms, a persistent form and an ephemeral (mutable) form; efficient conversion operations are offered. François Pottier subsequently implemented a test harness, based on afl-fuzz, which allowed us to submit Émilie's code to intensive testing and detect and fix a few bugs. This work is not yet published; we intend to pursue it in 2019, to publish the library and perhaps to verify it.

7.5.5. *TLA+*

Participants: Damien Doligez, Leslie Lamport [Microsoft Research], Ioannis Filippidis, 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 [36]. This requires building tools to help write TLA+ specifications and mechanically check proofs.

Since October 2018, Ioannis Filippidis has been working on extending the TLAPS tool to deal with proofs of temporal properties. Under some well-defined circumstances, an occurrence of the ENABLED operator applied to a formula f can be replaced by a version of f where the primed variables are replaced by new existentially-quantified variables. The result is a first-order formula that can be sent to one of TLAPS's first-order back-ends. This rewriting of ENABLED suffices to prove a large class of liveness properties. Ioannis has started implementing this in TLAPS.

MARELLE Project-Team

6. New Results

6.1. Extension language for Coq

Participants: Enrico Tassi, Feruccio Guidi [University of Bologna], Claudio Sacerdoti Coen [University of Bologna].

We continued our work on the design of a language mixing λ-prolog and constraint programming. This year, we redesigned and provided a new implementation of the constraint handling rules, leading to a first public release of the software. We are starting to have users beyond our own team:

- (Inria/Parsifal) MLTS https://github.com/voodoos/mlts
- (Inria/Parsifal) proofcert https://github.com/proofcert/checkers
- (UML.eu) Lang-n-play https://github.com/mcimini/lang-n-play

In an article submitted for publication [24], we showed that Elpi could be used to give a short implementation of Type Theory.

We are also starting a collaboration to construct an elaborator for HOL-Light using Elpi.

6.2. Deriving equality tests

Participant: Enrico Tassi.

In type theory, for most inductive types, it is possible to construct a two-argument boolean function that tests when two terms of the type are equal. When inductive types have constructors containing sub-components from another inductive, this needs to be done in a modular way. This year, we studied how this problem could be solved in a modular way using Elpi. It turns out that the unary parametricity translation can serve as a tool to make the derivation compositional. This is described in a pre-print [25].

6.3. Parametricity proofs

Participants: Cyril Cohen, Abishek Anand [Cornell University], Simon Boulier [Inria Gallinette], Matthieu Sozeau [Inria Pi.r2], Nicolas Tabareau [Inria Gallinette], Robert Y. Lewis [Vrije Universiteit Amsterdam], Johannes Hölzl [CMU, Pittsburgh, USA and Vrije Universiteit, Amsterdam, the Netherlands].

After our previous experiment using Elpi to develop a tool that produces parametricity proofs, we investigated the use of the *Template-Coq* framework to implement this kind of algorithm. This work is described in [11]. A similar experiment has been performed using the Lean theorem prover.

6.4. Proving Expected Sensitivity of Probabilistic Programs

Participants: Benjamin Grégoire, Gilles Barthe [IMDEA], Thomas Espitau [UPMC Paris 6], Justin Hsu [University of Pennsylvania], Pierre-Yves Strub [Ecole Polytechnique].

Program sensitivity, also known as Lipschitz continuity, describes how small changes in a program's input lead to bounded changes in the output. We propose an average notion of program sensitivity for probabilistic programs—expected sensitivity—that averages a distance function over a probabilistic coupling of two output distributions from two similar inputs. This work is described in [8].

6.5. An Assertion-Based Program Logic for Probabilistic Programs

Participants: Benjamin Grégoire, Gilles Barthe [IMDEA], Thomas Espitau [UPMC Paris 6], Marco Gaboardi [University at Buffalo, SUNY], Justin Hsu [University of Pennsylvania], Pierre-Yves Strub [Ecole Polytechnique].

We have developed Ellora, a sound and relatively complete assertion-based program logic, and demonstrate its expressivity by verifying several classical examples of randomized algorithms using an implementation in the EasyCrypt proof assistant. Ellora features new proof rules for loops and adversarial code, and supports richer assertions than existing program logics. We also show that Ellora allows convenient reasoning about complex probabilistic concepts by developing a new program logic for probabilistic independence and distribution law, and then smoothly embedding it into Ellora. This is described in article [14].

6.6. Vectorizing Higher-Order Masking

Participants: Benjamin Grégoire, Kostas Papagiannopoulos [Radboud University], Peter Schwabe [Radboud University], Ko Stoffelen [Radboud University].

The cost of higher-order masking as a countermeasure against side-channel attacks is often considered too high for practical scenarios, as protected implementations become very slow. At Eurocrypt 2017, we have proposed the bounded moment leakage model to study the (theoretical) security of parallel implementations of masking schemes. In this work we show how the NEON vector instructions of larger ARM Cortex-A processors can be exploited to build much faster masked implementations of AES based on the bounded moment model. This work is described in publication [18].

6.7. Masking the GLP Lattice-Based Signature Scheme at Any Order

Participants: Benjamin Grégoire, Gilles Barthe [IMDEA], Sonia Belaïd [CryptoExpert], Thomas Espitau [UPMC Paris 6], Pierre-Alain Fouque [Université Rennes 1], Mélissa Rossi [ENS Paris], Mehdi Tibouchi [NTT].

Recently, numerous physical attacks have been demonstrated against lattice based schemes, often exploiting their unique properties such as the reliance on Gaussian distributions, rejection sampling and FFT-based polynomial multiplication. In this work, we describe the first masked implementation of a lattice-based signature scheme. Since masking Gaussian sampling and other procedures involving contrived probability distribution would be prohibitively inefficient, we focus on the GLP scheme. This work is described in [13].

6.8. Symbolic Proofs for Lattice-Based Cryptography

Participants: Benjamin Grégoire, Gilles Barthe [IMDEA], Xiong Fan [Cornell], Joshua Gancher [Cornell], Charlie Jacomme [LSV], Elaine Shi [Cornell].

Symbolic methods have been used extensively for proving security of cryptographic protocols in the Dolev-Yao model, and more recently for proving security of cryptographic primitives and constructions in the computational model. However, existing methods for proving security of cryptographic constructions in the computational model often require significant expertise and interaction, or are fairly limited in scope and expressivity. In this work we introduce a symbolic approach for proving security of cryptographic constructions based on the Learning With Errors assumption. This work is described in [15].

6.9. Formal Security Proof of CMAC and Its Variants

Participants: Benjamin Grégoire, Cécile Baritel-Ruet, François Dupressoir [University of Surrey], Pierre-Alain Fouque [Université Rennes 1].

The CMAC standard, when initially proposed by Iwata and Kurosawa as OMAC1, was equipped with a complex game-based security proof. Following recent advances in formal verification for game-based security proofs, we have formalized a proof of unforgeability for CMAC in EasyCrypt. This work is described in [12].

6.10. Secure Compilation of Side-Channel Countermeasures: The Case of Cryptographic "Constant-Time"

Participants: Benjamin Grégoire, Gilles Barthe [IMDEA], Vincent Laporte [IMDEA].

Software-based countermeasures provide effective mitigation against side-channel attacks, often with minimal efficiency and deployment overheads. Their effectiveness is often amenable to rigorous analysis: specifically, several popular countermeasures can be formalized as information flow policies, and correct implementation of the countermeasures can be verified with state-of-the-art analysis and verification techniques. However, in absence of further justification, the guarantees only hold for the language (source, target, or intermediate representation) on which the analysis is performed. We consider the problem of preserving side-channel counter-measures by compilation for cryptographic "constant-time", a popular countermeasure against cache-based timing attacks. We have presented a general method, based on the notion of constant-time-simulation, for proving that a compilation pass preserves the constant-time countermeasure. This work was described in [16]. At the conference, this work received the "distinguished paper" award.

6.11. Hypotheses of Decisional Diffie-Hellmann

Participants: Benjamin Grégoire, Mohamad El Laz, Tamara Rezk [Inria, Indes project team].

In the thesis work of Mohamad El Laz, co-supervised by Benjamin Grégoire and Tamara Rezk (Indes project-team), we studied the cryptographic hypothesis of DDH (Decisional Diffie-Hellman) and implementations that would break this hypothesis. We focused on ElGamal encryption cryptosystem implementations to assess they use the DDH hypothesis correctly. We analyzed a number of implementations including Botan, Belenios and Libgcrypt. The lessons learned from this analysis are that the hypotheses are not always well understood.

In a second stage we considered message encoding methods. We investigated several approaches such as DCDH (Decisional Class Diffie-Hellman) in Encoding-Free ElGamal Encryption.

6.12. Proving the domain management protocol

Participants: José Bacelar Almeida [INESC TEC], Manuel Barbosa [INESC TEC], Gilles Barthe [IMDEA], Benjamin Grégoire, Vitor Pereira [INESC TEC], Bernardo Portela [INESC TEC], Benedikt Schmidt [Google Inc.], François-Xavier Standaert [Université Catholique de Louvain], Pierre-Yves Strub [Ecole Polytechnique].

We have performed a machine-checked proof of security for the domain management protocol of Amazon Web Services KMS (Key Management Service), a critical security service used throughout AWS and by AWS customers. Domain management is at the core of KMS; it governs the long-term keys that anchor the security of encryption services at AWS. Informally, we show that the protocol securely implements a distributed encryption mechanism. Formally, the proof shows that the domain management protocol is indistinguishable from an ideal encryption functionality under standard cryptographic assumptions.

6.13. Formalized graph theory algorithms

Participants: Cyril Cohen, Laurent Théry, Ran Chen [Chinese Academy of Science], Jean-Jacques Lévy [Inria Pi.r2], Stephan Merz [Inria Veridis].

We formalise the correctness proof of Tarjan's algorithm for computing strongly connected components using the Mathematical Component Library. This leads to a comparison of formalisation between various systems described in [22].

6.14. Formal study of a triangulation algorithm

Participant: Yves Bertot.

In work from 2010, a formal description of Delaunay triangulations was presented where the input was a triangulation not satisfying the Delaunay criterion and where the output was a triangulation satisfying this criterion.

In this work, we wish to complete the previous work by describing an algorithm that produces the initial triangulation. We plan this work in several phases, where the first phase only uses simple data-structures, more advanced structures being introduced only later. This work was presented partially in an invited talk at the ICTAC conference [10].

6.15. 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 implemented a paper of Sierpinski about properties of continuous ordinal functions and limits of such functions.

We implemented a paper on sums of sequences of ordinals, showing that the value obtained (which depends on the order) lies in a finite set. We also showed that this result does not hold when replacing ordinals by order types.

We implemented a paper by Tarski that says if every infinite cartinal is equal to its square, then every set can be well-ordered (this is the axiom of choice). We had to modify our library to make the use of the axiom of choice more explicit.

We continued implementing in Coq the Exercises of Set Theory of Bourbaki. We solved two of them, and proved by a counter example that three of them are false.

6.16. Formal study of double-word arithmetic algorithms

Participants: Laurence Rideau, Jean-Michel Muller [CNRS and ENS Lyon], Valentina Popescu [CNRS and ENS Lyon], Mioara Joldes [CNRS LAAS].

As part of the ANR Fastrelax project, we are formalizing 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" [27]. The formalization is progressing, moving from addition to multiplication. The progress is slowed down because minor errors in the informal proofs are regularly uncovered, which requires a dialog with the initial authors.

6.17. Proofs of transcendence

Participants: Sophie Bernard, Yves Bertot, Laurence Rideau.

The work on proofs of transcendence that was started the previous year was completed this year by an effort to integrate generic part of the proofs in the Mathematical Components library. A public package for easy re-use by other researchers was also developed.

6.18. Abel's theorem

Participants: Sophie Bernard, Yves Bertot, Cyril Cohen, Laurence Rideau, Assia Mahboubi [Inria Gallinette], Russell O'Connor [McMaster University].

A natural extension of the work on group theory is a proof that polynomials of degree higher than 5 cannot be solved by radicals. This is known as Abel's theorem. We have started an experiment to give a formal proof of this result on top of the Mathematical Components library.

6.19. Formalizing Hermitian Forms

Participants: Cyril Cohen, Laurence Rideau.

We updated the representation and relevant theorems for bilinear, sesquilinear, and hermitian forms in the Mathematical Components library and updated the archived proof of the odd-order theorem (Feit-Thompson) to use the new presentation. This work also includes a proof of the Spectral Theorem.

6.20. Mathematical Components Analysis

Participants: Cyril Cohen, Damien Rouhling, Reynald Affeldt [AIST Japan], Assia Mahboubi [Inria Gallinette], Pierre-Yves Strub [Ecole Polytechnique].

As a synthesis of the lessons learned in the usage of Mathematical Components and Coquelicot, we develop an extension of the Mathematical Components library to cover questions of analysis. This work includes a new tactic called `near` to handle reasoning steps around limits and filters and little-o notation (following Landau's style of asymptotic reasoning). This work is described in [6]. There also contains a new formalization of topoligical structures, Rolle's theorem, the intermediate value theorem, and Heine Borel's theorem. Ongoing work concentrates on a better design of the topological hierarchy and a simplification of the properties expected from real numbers (following a design by A. Mahboubi and P.-Y. Strub).

Some of this work also includes experiments performed with the LEAN theorem prover (developed at Microsoft Research).

6.21. Rigorous Polynomial Approximation

Participants: Florian Steinberg, Laurent Théry.

We have developed a certified library for computing Chebyshev models for formulas composed of polynomials, exponential, logarithm, and trigonometric function. This work is part of the ANR project FastRelax. The code is available at https://github.com/FlorianSteinberg/Cheby

6.22. Formalization of proofs in control theory

Participants: Damien Rouhling, Cyril Cohen.

Damien Rouhling presented his work on formalizing control theory for an inverted pendulum at an international conference in January [19].

The original development was based on Coquelicot. An analysis of the difficulties in formalizing led to the design of Mathematical Components Analysis. The development on control was then ported to this new library. This work was presented at the Coq Workshop in July.

6.23. 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. An article on this topic was published [17]. This is also the one of the main topics of Boris Djalal's thesis, which was defended in December.

6.24. A type theory for Algebraic Structures

Participants: Cyril Cohen, Assia Mahboubi, Xavier Montillet.

In collaboration with members of the Inria Gallinette team, we are investigating the properties that a type theory should enjoy to support algebraic structures better than what is currently available.

MEXICO Project-Team

7. New Results

7.1. Contract Based Design of Symbolic Controllers for Interconnected Multiperiodic Sampled-Data Systems

This paper deals with the synthesis of symbolic controllers for interconnected sampled-data systems where each component has its own sampling period. A compositional approach based on continuous-time assume-guarantee contracts is used. We provide sufficient conditions guaranteeing for a sampled-data system, satisfaction of an assume-guarantee contract and completeness of trajectories. Then, compositional results can be used to reason about interconnection of mul-tiperiodic sampled-data systems. We then show how discrete abstractions and symbolic control techniques can be applied to enforce the satisfaction of contracts and ensure completeness of trajectories. Finally, theoretical results are applied to a vehicle platooning problem on a circular road, which show the effectiveness of our approach.

7.2. Boolean Networks: Beyond Generalized Asynchronicity

Boolean networks are commonly used in systems biology to model dynamics of biochemical networks by abstracting away many (and often unknown) parameters related to speed and species activity thresholds. It is then expected that Boolean networks produce an over-approximation of behaviours (reachable configurations), and that subsequent refinements would only prune some impossible transitions. However, we show that even generalized asynchronous updating of Boolean networks, which subsumes the usual updating modes including synchronous and fully asynchronous, does not capture all transitions doable in a multi-valued or timed refinement. We define a structural model transformation which takes a Boolean network as input and outputs a new Boolean network whose asynchronous updating simulates both synchronous and asynchronous updating of the original network, and exhibits even more behaviours than the generalized asynchronous updating. We argue that these new behaviours should not be ignored when analyzing Boolean networks, unless some knowledge about the characteristics of the system explicitly allows one to restrict its behaviour.

7.3. Most Permissive Semantics of Boolean Networks

The usual update modes of Boolean networks (BNs), including synchronous and (generalized) asynchronous, fail to capture behaviours introduced by multivalued refinements. Thus, update modes do not allow a correct abstract reasoning on dynamics of biological systems, as they may lead to reject valid BN models. We introduce a new semantics for interpreting BNs which meets with a correct abstraction of any multivalued refinements, with any update mode. This semantics subsumes all the usual updating modes, while enabling new behaviours achievable by more concrete models. Moreover, it appears that classical dynamical analyses of reachability and attractors have a simpler computational complexity: – reachability can be assessed in a polynomial number of iterations (instead of being PSPACE-complete with update modes); – attractors are hypercubes, and deciding the existence of attractors with a given upper-bounded dimension is in NP (instead of PSPACE-complete with update modes). The computation of iterations is in NP in the very general case, and is linear when local functions are monotonic, or with some usual representations of functions of BNs (binary decision diagrams, Petri nets, automata networks, etc.). In brief, the most permissive semantics of BNs enables a correct abstract reasoning on dynamics of BNs, with a greater tractability than previously introduced update modes. This technical report lists the main definitions and properties of the most permissive semantics of BNs, and draw some remaining open questions.

7.4. Concurrency in Boolean networks

Boolean networks (BNs) are widely used to model the qualitative dynamics of biological systems. Besides the logical rules determining the evolution of each component with respect to the state of its regulators, the scheduling of components updates can have a dramatic impact on the predicted behaviours. In this paper, we explore the use of Contextual Petri Nets (CPNs) to study dynamics of BNs with a concurrency theory perspective. After showing bi-directional translations between CPNs and BNs and analogies between results on synchronism sensitivities, we illustrate that usual updating modes for BNs can miss plausible behaviours, i.e., incorrectly conclude on the absence/impossibility of reaching specific configurations. Taking advantage of CPN semantics enabling more behaviour than the generalized asynchronous updating mode, we propose an encoding of BNs ensuring a correct abstraction of any multivalued refinement, as one may expect to achieve when modelling biological systems with no assumption on its time features.

7.5. On the Composition of Discrete and Continuous-time Assume-Guarantee Contracts for Invariance

Many techniques for verifying invariance properties are limited to systems of moderate size. In this paper, we propose an approach based on assume-guarantee contracts and compositional reasoning for verifying invariance properties of a broad class of discrete-time and continuous-time systems consisting of interconnected components. The notion of assume-guarantee contracts makes it possible to divide responsibilities among the system components: a contract specifies an invariance property that a component must fulfill under some assumptions on the behavior of its environment (i.e. of the other components). We define weak and strong semantics of assume-guarantee contracts for both discrete-time and continuous-time systems. We then establish a certain number of results for compositional reasoning, which allow us to show that a global invariance property of the whole system is satisfied when all components satisfy their own contract. Interestingly, we show that the weak satisfaction of the contract is sufficient to deal with cascade compositions, while strong satisfaction is needed to reason about feedback composition. Specific results for systems described by differential inclusions are then developed. Throughout the paper, the main results are illustrated using simple examples.

7.6. Compositional synthesis of state-dependent switching control

We present a correct-by-design method of state-dependent control synthesis for sampled switching systems. Given a target region R of the state space, our method builds a capture set S and a control that steers any element of S into R. The method works by iterated backward reachability from R. The method is also used to synthesize a recurrence control that makes any state of R return to R infinitely often. We explain how the synthesis method can be performed in a compositional manner, and apply it to the synthesis of a compositional control of a concrete floor-heating system with 11 rooms and up to $2^1 1 = 2048$ toswitching modes.

7.7. An Improved Algorithm for the Co3ntrol Synthesis of Nonlinear Sampled Switched Systems

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.

7.8. Control Synthesis for Stochastic Switched Systems using the Tamed Euler Method

In this paper, we explain how, under the one-sided Lipschitz (OSL) hypothesis, one can find an error bound for a variant of the Euler-Maruyama approximation method for stochastic switched systems. We then explain how this bound can be used to control stochastic switched switched system in order to stabilize them in a given region. The method is illustrated on several examples of the literature.

7.9. The Complexity of Diagnosability and Opacity Verification for Petri Nets

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.

7.10. Integrating Simulink Models into the Model Checker Cosmos

We present an implementation for Simulink model executions in the statistical model-checker Cosmos. We take profit of this implementation for an hybrid modeling combining Petri nets and Simulink models.,Nous présentons une implémentation pour l'exécution de modèles Simulink dans le model-checker Cosmos. Cette implémentation est ensuite utilisée pour la simulation de modèles hybrides, combinant des réseaux de Petri et des modèles Simulink.

7.11. Bounds Computation for Symmetric Nets

Monotonicity in Markov chains is the starting point for quantitative abstraction of complex probabilistic systems leading to (upper or lower) bounds for probabilities and mean values relevant to their analysis. While numerous case studies exist in the literature, there is no generic model for which monotonicity is directly derived from its structure. Here we propose such a model and formalize it as a subclass of Stochastic Symmetric (Petri) Nets (SSNs) called Stochastic Monotonic SNs (SMSNs). On this subclass the monotonicity is proven by coupling arguments that can be applied on an abstract description of the state (symbolic marking). Our class includes both process synchronizations and resource sharings and can be extended to model open or cyclic closed systems. Automatic methods for transforming a non monotonic system into a monotonic one matching the MSN pattern, or for transforming a monotonic system with large state space into one with reduced state space are presented. We illustrate the interest of the proposed method by expressing standard monotonic models and modelling a flexible manufacturing system case study.

7.12. Distributed computation of vector clocks in Petri nets unfolding for test selection

Petri net unfoldings with time stamps allow to build distributed testers for distributed systems. However, the construction of the annotated unfolding of a distributed system currently remains a centralized task. In the aforemention paper, we extend a distributed unfolding technique in order to annotate the resulting unfolding with time stamps. This allows for distributed construction of distributed testers for distributed systems.

7.13. Hyper Partial Order Logic

We define HyPOL, a local hyper logic for partial order models, expressing properties of sets of runs. These properties depict shapes of causal dependencies in sets of partially ordered executions,with similarity relations defined as isomorphisms of past observations. Unsurprisingly, since comparison of projections are included, satisfiability of this logic is undecidable. We then addressmodel checking of HyPOL and show that, already for safe Petri nets, the problem is undecidable. Fortunately, sensible restrictions of observations and nets allow us to bring back model checking ofHyPOL to a decidable problem, namely model checking of MSO on graphs of bounded treewidth.

7.14. Integrating Simulink Models into the Model Checker Cosmos

We present an implementation for Simulink model executions in the statistical model-checker Cosmos. We take profit of this implementation for hybrid modeling and simulations combining Petri nets and Simulink models.

7.15. Site-Directed Deletion

We introduce a new bio-inspired operation called a site-directed deletion motivated from site-directed mutagenesis performed by enzymatic activity of DNA polymerase: Given two strings x and y, a site-directed deletion partially deletes a substring of x guided by the string y that specifies which part of a substring can be deleted. We study a few decision problems with respect to the new operation and examine the closure properties of the (iterated) site-directed deletion operations. We, then, define a site-directed deletion-closed (and-free) language L and investigate its decidability properties when L is regular or context-free.

7.16. Site-Directed Insertion: Decision Problems, Maximality and Minimality

Site-directed insertion is an overlapping insertion operation that can be viewed as analogous to the overlap assembly or chop operations that concatenate strings by overlapping a suffix and a prefix of the argument strings. We consider decision problems and language equations involving site-directed insertion. By relying on the tools provided by semantic shuffle on trajectories we show that one variable equations involving site-directed insertion and regular constants can be solved. We consider also maximal and minimal variants of the site-directed insertion operation.

7.17. A Faithful Binary Circuit Model with Adversarial Noise

Accurate delay models are important for static and dynamic timing analysis of digital circuits, and mandatory for formal verification. However, Függer et al. [IEEE TC 2016] proved that pure and inertial delays, which are employed for dynamic timing analysis in state-of-the-art tools like ModelSim, NC-Sim and VCS, do not yield faithful digital circuit models. Involution delays, which are based on delay functions that are mathematical involutions depending on the previous-output-to-input time offset, were introduced by Függer et al. [DATE'15] as a faithful alternative (that can easily be used with existing tools). Although involution delays were shown to predict real signal traces reasonably accurately, any model with a deterministic delay function is naturally limited in its modeling power. In this paper, we thus extend the involution model, by adding non-deterministic delay variations (random or even adversarial), and prove analytically that faithfulness is not impaired by this generalization. Albeit the amount of non-determinism must be considerably restricted to ensure this property, the result is surprising: the involution model differs from non-faithful models mainly in handling fast glitch trains, where small delay shifts have large effects. This originally suggested that adding even small variations should break the faithfulness of the model, which turned out not to be the case. Moreover, the results of our simulations also confirm that this generalized involution model has larger modeling power and, hence, applicability.

7.18. Tight Bounds for Asymptotic and Approximate Consensus

We study the performance of asymptotic and approximate consensus algorithms under harsh environmental conditions. The asymptotic consensus problem requires a set of agents to repeatedly set their outputs such that the outputs converge to a common value within the convex hull of initial values. This problem, and the related approximate consensus problem, are fundamental building blocks in distributed systems where exact consensus among agents is not required or possible, e.g., man-made distributed control systems , and have applications in the analysis of natural distributed systems, such as flocking and opinion dynamics. We prove tight lower bounds on the contraction rates of asymptotic consensus algorithms in dynamic networks, from which we deduce bounds on the time complexity of approximate consensus algorithms. In particular, the obtained bounds show optimality of asymptotic and approximate consensus algorithms presented in [Charron-Bost et al., ICALP'16] for certain dynamic networks, including the weakest dynamic network model in which asymptotic and approximate consensus are solvable. As a corollary we also obtain asymptotically tight bounds for asymptotic consensus in the classical asynchronous model with crashes. Central to our lower bound proofs is an extended notion of valency, the set of reachable limits of an asymptotic consensus algorithm starting from a given configuration. We further relate topological properties of valencies to the solvability of exact consensus , shedding some light on the relation of these three fundamental problems in dynamic networks.

7.19. Pomsets and Unfolding of Reset Petri Nets

Reset Petri nets are a particular class of Petri nets where transition firings can remove all tokens from a place without checking if this place actually holds tokens or not. In this paper we look at partial order semantics of such nets. In particular, we propose a pomset bisimulation for comparing their concurrent behaviours. Building on this pomset bisimulation we then propose a generalization of the standard finite complete prefixes of unfolding to the class of safe reset Petri nets.

7.20. Fast All-Digital Clock Frequency Adaptation Circuit for Voltage Droop Tolerance

Naive handling of supply voltage droops in synchronous circuits results in conservative bounds on clock speeds, resulting in poor performance even if droops are rare. Adaptive strategies detect such potentially hazardous events and either initiate a rollback to a previous state or proactively reduce clock speed in order to prevent timing violations. The performance of such solutions critically depends on a very fast response to droops. However, state-of-the-art solutions incur synchronization delay to avoid that the clock signal is affected by metastability. Addressing the challenges discussed by Keith Bowman in his ASYNC 2017 keynote talk, we present an all-digital circuit that can respond to droops within a fraction of a clock cycle. This is achieved by delaying clock signals based on measurement values while they undergo synchronization simultaneously. We verify our solution by formally proving correctness, complemented by VHDL and Spice simulations of a 65 nm ASIC design confirming the theoretically obtained results.

7.21. Fast Multidimensional Asymptotic and Approximate Consensus

We study the problems of asymptotic and approximate consensus in which agents have to get their values arbitrarily close to each others' inside the convex hull of initial values, either without or with an explicit decision by the agents. In particular, we are concerned with the case of multidimensional data, i.e., the agents' values are d-dimensional vectors. We introduce two new algorithms for dynamic networks, subsuming classical failure models like asynchronous message passing systems with Byzantine agents. The algorithms are the first to have a contraction rate and time complexity independent of the dimension d. In particular, we improve the time complexity from the previously fastest approximate consensus algorithm in asynchronous message passing systems with Byzantine faults by Mendes et al. [Distrib. Comput. 28].

7.22. Parameter Space Abstraction and Unfolding Semantics of Discrete Regulatory Networks

The modelling of discrete regulatory networks combines a graph specifying the pairwise influences between the variables of the system, and a parametrisation from which can be derived a discrete transition system. Given the influence graph only, the exploration of admissible parametrisations and the behaviours they enable is computationally demanding due to the combinatorial explosions of both parametrisation and reachable state space. This article introduces an abstraction of the parametrisation space and its refinement to account for the existence of given transitions, and for constraints on the sign and observability of influences. The abstraction uses a convex sub-lattice containing the concrete parametrisation space specified by its infimum and supremum parametrisations. It is shown that the computed abstractions are optimal, i.e., no smaller convex sublattice exists. Although the abstraction may introduce over-approximation, it has been proven to be conservative with respect to reachability of states. Then, an unfolding semantics for Parametric Regulatory Networks is defined, taking advantage of concurrency between transitions to provide a compact representation of reachable transitions. A prototype implementation is provided: it has been applied to several examples of Boolean and multi-valued networks, showing its tractability for networks with numerous components.

7.23. 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.

7.24. Diagnosability of Repairable Faults

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. 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.

7.25. Metastability-Containing Circuits

In digital circuits, metastability can cause deteriorated signals that neither are logical 0 nor logical 1, breaking the abstraction of Boolean logic. Synchronizers, the only traditional countermeasure, exponentially decrease the odds of maintained metastability over time. We propose a fundamentally different approach: It is possible to deterministically contain metastability by fine-grained logical masking so that it cannot infect the entire circuit. At the heart of our approach lies a time-and value-discrete model for metastability in synchronous clocked digital circuits, in which metastability is propagated in a worst-case fashion. The proposed model permits positive results and passes the test of reproducing Marino's impossibility results. We fully classify which functions can be computed by circuits with standard registers. Regarding masking registers, we show that more functions become computable with each clock cycle, and that masking registers permit exponentially smaller circuits for some tasks. Demonstrating the applicability of our approach, we present the first fault-tolerant distributed clock synchronization algorithm that deterministically guarantees correct behavior in the presence of metastability. As a consequence, clock domains can be synchronized without using synchronizers, enabling metastability-free communication between them.

MOCQUA Team

7. New Results

7.1. Completeness of the ZX-calculus

- Participants: Renaud Vilmart, Simon Perdrix, Emmanuel Jeandel

The ZX-Calculus is a powerful graphical language for quantum reasoning and quantum computing introduced by Bob Coecke and Ross Duncan [36]. The ZX-calculus has several applications in quantum information processing [37] (e.g. measurement-based quantum computing, quantum codes, foundations), and can be used through the interactive theorem prover Quantomatic. However, the main obstacle to wider use of the ZX-calculus was the absence of a *completeness* result for a *universal* fragment of quantum mechanics, in order to guarantee that any true property is provable using the ZX-calculus. We have introduced the first complete axiomatisation for a universal fragment of quantum mechanics. We also showed that a single additional rule makes the ZX-calculus complete for the whole pure qubit quantum mechanics. These results have been presented at LICS this year [16], [17] and will be presented at QIP'19, the main conference in quantum information processing.

7.2. Second-order entropy accumulation theorem

- Participants: Frédéric Dupuis

Device-independent cryptography is a way to use quantum mechanics to perform cryptographic tasks using equipment from an untrusted manufacturer. To prove the security of device-independent protocols, the main challenge is to show that a step-by-step procedure involving the untrusted device produces a certain of randomness even from the point of view of the manufacturer. The entropy accumulation theorem [38] provides a generic way to obtain such statements. However, while the bounds provided by this theorem are optimal in the first order (meaning the term that is linear in the number of steps in the process), the second-order sublinear term is bounded more crudely, in such a way that the bounds deteriorate significantly when the theorem is applied directly to protocols where parameter estimation is done by sampling a small fraction of the positions, as is done in most QKD protocols. In [25], we improve this second-order sublinear term and remedy this problem. This paper has been submitted to IEEE Transactions on Information Theory.

7.3. Mixed-state certification

- Participants: Frédéric Dupuis

Mixed-state certification consists of ensuring that a quantum state on n subsystems is close to n copies of a given mixed state, up to a small number of errors, by sampling a small fraction of the positions. While this task makes no sense classically (it effectively amounts to certifying that a string came from a particular probability distribution), it makes sense quantumly if we can ask someone (that we call a prover) to supply purifications of the sampled positions. However, such sampling procedures cannot be analyzed straightforwardly using standard sampling results, and care must be taken even when defining what success means. In [26], we introduced these concepts, and we showed that this sampling protocol offers secure certification in the presence of a possibly dishonest prover. We then applied this result to two-party quantum coin-tossing. This work was presented at QCrypt 2018 and TCC 2018 (and will appear in the proceedings of the latter).

7.4. Descriptive Set Theory

- Participants: Mathieu Hoyrup

Descriptive Set Theory (DST) aims at classifying sets and functions in terms of the complexity of describing them. It is closely related to logic and computation theory, where sets and functions can be described by logical formulas or computer programs. DST was originally developed on a restricted class of topological spaces, the Polish spaces, which does not cover important classes of spaces that are needed in Theoretical Computer Science, especially in programming semantics, notably (Scott) domains or spaces of higher-order (Kleene-Kreisel) functionals. We investigate DST on such spaces and show that it does not work as nicely as on usual spaces. The article [29] is currently submitted. This work has been presented during an invited talk at CiE 2018 [13].

7.5. Semicomputable geometry

- Participants: Mathieu Hoyrup

Semicomputability is a natural notion arising from logic and theoretical computer science. Termination of programs is not decidable but semidecidable. Semicomputability of subsets of the plane is an important notion. For instance whether the famous Mandelbrot set is computable is still an open problem, while its semicomputability is easy to prove. Intuitively, we can write a program that progressively fills out the complement of the set, but we do not know when the picture is complete. We studied semicomputability of much simpler sets, namely filled triangles. While this problem looks simple at first sight, it is considerably rich and raises many questions. What properties should the coordinates of the vertices of a triangle satisfy to make it semicomputable? How can we parametrize such triangles? What happens for other sets such as disks or general convex sets? We developed a thorough study of these problems in [15].

7.6. Resource bounded computation

- Participants: Emmanuel Hainry

Controlling resource consumption is a crucial aspect of programming. Resources such as time, space, intrication are limited, and helping the programmer to avoid overconsumption or pointing problematic code is an important endeavor. We introduced a type-system for an Object Oriented Programming Language (*à la* Java) that gives a guarantee of polynomial-time computability provided that the program halts [12]. This result has several interesting features as it works with complex object data-structures in a real-like programming language; checking the type system is polynomial time decidable; we provided a O bound hence giving an explicit worst case complexity bound.

7.7. Inductive reasoning

- Participants: Isabelle Gnaedig, Sofien Ben Ayed

We are interested in quantifying the power of axiomatic theories. For this purpose, induction is a key concept. We have investigated the different validity proofs of inductive reasoning, the equivalence of induction with the well-ordered principle and well-foundedness, the differences between first and second order forms of the induction principle, and the notion of ω-consistency, qualifying theories interpreting arithmetic for which proving a property for each value of standard integers does not imply that the property is always true. We have also studied the importance of the axiom of choice for induction, and analysed a recent interpretation of induction by Hardin and Taylor through the hat problem [22].

7.8. Cellular automata with stochastic evolutions

- Participants: Nazim Fatès, Irène Marcovici

In order to explore the computing abilities of simple stochastic cellular automata, we tackle the case of Alesia, a two-player zero-sum game which is quite similar to the rock-paper-scissors game. In this game, two players simultaneously move and do not know what the opponent plays at a given round. The simultaneity of the moves implies that there is no deterministic good strategy in this game, otherwise one would anticipate the moves of the opponent and easily win the game. We explored how to build a family of one-dimensional stochastic cellular automata to play this game by progressively increasing the complexity of the transitions. We showed the possibility to construct a family of rules with interesting results, including good performance when confronted to the Nash-equilibrium strategy [14].

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 (IIT Karaghpur, India), we studied a particular form of reversibility: the possibility of returning infinitely often to the initial condition after a random number of time steps. This is the recurrence property of the system. We analyzed this property for the simple rules and described the communication graph of the system [21].

We also contributed to the diffusion of some already-established knowledge on the simulation of complex systems in Biology, more precisely in the case of the formation of swarms [19] and in the case of asynchronous cellular automata [20].

PARSIFAL Project-Team

7. New Results

7.1. Functional programming with λ-tree syntax

Participants: Ulysse Gerard, Dale Miller, Gabriel Scherer.

We have been designing a new functional programming language, MLTS, that uses the λ-*tree* syntax approach to encoding bindings that appear within data structures [17]. In this setting, bindings never become free nor escape their scope: instead, binders in data structures are permitted to *move* into binders within programs phrases. The design of MLTS—whose concrete syntax is based on that of OCaml—includes additional sites within programs that directly support this movement of bindings. Our description of MLTS includes a typing discipline that naturally extends the typing of OCaml programs.

The operational semantics of MLTS is given using natural semantics for evaluation. We shall view such natural semantics as a logical theory with a rich logic that includes both nominal abstraction and the ∇-quantifier: as a result, the natural semantic specification of MLTS can be given a succinct and elegant presentation.

We have developed a number of examples of how this new programming language can be used. Some of the most convincing of these examples are programs that manipuate untyped λ-terms. A web-based implementation of an MLTS interpreter is available to anyone with a modern web browser: simply visit https://trymlts.github.io/. Small MLTS programs can be composed and executed using that interpreter.

7.2. Proof theory for model checking

Participant: Dale Miller.

While model checking has often been considered as a practical alternative to building formal proofs, we have argued that the theory of sequent calculus proofs can be used to provide an appealing foundation for model checking [7]. Given that the emphasis of model checking is on establishing the truth of a property in a model, our framework concentrates on *additive* inference rules since these provide a natural description of truth values via inference rules. Unfortunately, using these rules alone can force the use of inference rules with an infinite number of premises. In order to accommodate more expressive and finitary inference rules, *multiplicative* rules must be used, but limited to the construction of *additive synthetic inference rules*: such synthetic rules are described using the proof-theoretic notions of polarization and focused proof systems. This framework provides a natural, proof-theoretic treatment of reachability and non-reachability problems, as well as tabled deduction, bisimulation, and winning strategies. (Q. Heath collaborated on several parts of this research effort.)

7.3. From syntactic proofs to combinatorial proofs

Participants: Matteo Acclavio, Lutz Straßburger.

We continued our research on combinatorial proofs as a notion of proof identity for classical logic. We managed to extend our results from last year: We show for various syntactic formalisms including sequent calculus, analytic tableaux, and resolution, how they can be translated into combinatorial proofs, and which notion of identity they enforce. This allows the comparison of proofs that are given in different formalisms.

These results have been presented at the MLA workshop ins Kanazawa and the IJCAR conference in Oxford, published in [25].

7.4. Proof nets for first-order additive linear logic

Participant: Lutz Straßburger.

In a joint work with Willem Heijltjes (University of Bath) and Dominic Hughes (UC Berkeley) we present canonical proof nets for first-order additive linear logic, the fragment of linear logic with sum, product, and first-order universal and existential quantification. We present two versions of our proof nets. One, witness nets, retains explicit witnessing information to existential quantification. For the other, unification nets, this information is absent but can be reconstructed through unification. Unification nets embody a central contribution of the paper: first-order witness information can be left implicit, and reconstructed as needed. Witness nets are canonical for first-order additive sequent calculus. Unification nets in addition factor out any inessential choice for existential witnesses. Both notions of proof net are defined through coalescence, an additive counterpart to multiplicative contractibility, and for witness nets an additional geometric correctness criterion is provided. Both capture sequent calculus cut-elimination as a one-step global composition operation.

These results are published in [26] and have been presented at the First workshop of the Proof Society in Ghent and at the 3rd FISP workshop in Vienna.

7.5. On the Decision Problem for MELL

Participant: Lutz Straßburger.

The decision problem for multiplicative exponential linear logic (MELL) is one of the most important open problems in the are of linear logic. in 2015 there has been an attempt by Bimbò to prove the decidability of MELL. However, we have found several mistakes in that work, and the main mistake is so serious that there is no obvious fix, and therefore the decidability of MELL remains to be open. As a side effect, our work contains a complete (syntactic) proof of the decidability of the relevant version of MELL, that is the logic obtained from MELL by replacing the linear logic contraction rule by a general unrestricted version of the contraction rule. These results are presented in [27].

7.6. OCaml metatheory

Participant: Gabriel Scherer.

We worked on the evolution of advanced features of the OCaml programming language, designing static analyses to ensure their safety through a scientific study their metatheory. Specifically, we worked on unboxed type declarations (during an internship by Simon Colin, M1 from École Polytechnique) and recursive value definitions (during an internship by Alban Reynaud, L3 from ENS Lyon). The two internships and followup work each resulted in both a change proposal to the OCaml implementation and a submission to an academic conference.

7.7. Merlin: understanding a language server

Participant: Gabriel Scherer.

Thomas Réfis (Jane Street) and Frédéric Bour maintain the Merlin language server of OCaml, a tool that provides language-aware features to text editors. We collaborated with them on dissecting the tool and explaining its design and evolution ([4]); the similarities and differences with usual compiler frontends may inform future language implementation work, and our language-agnostic presentation may be of use to tool designers for other languages and proof assistants.

7.8. Language interoperability: ML and a Linear language

Participant: Gabriel Scherer.

In a programming system where programs are created in one programming language, we consider the addition of another programming language that interoperates with the first – and the reimplementation of some library/system functions in this new language. This can increase expressivity, but it could also break some assumptions made by programmers. Typically, adding a bridge to C or assembly code can introduce memory-unsafe code in a previously-safe system. In [18], we formalize a notion of "graceful" interoperability between two languages in this setting, determined by full abstraction, that is, preservation of equational reasoning. We instantiate this general idea by extending ML with an advanced expert language with linear types and linear mutable cells.

7.9. First-class simultaneous substitutions in the two-level logic approach

Participant: Kaustuv Chaudhuri.

The *two-level logic approach* that underlies the Abella prover is excellent at reasoning about the inductive structure of terms with binding constructs, such as λ-terms from the λ-calculus. However, there is no built in support in Abella for reasoning about the inductive structure of (simultaneous) substitutions. This lack of this kind of support is often criticized in the λ-tree syntax representational style that is used in Abella; indeed, in a number of other systems based on this style, support for reasoning about substitutions is explicitly added into the trusted kernel. In [14] we show how to formalize substitutions in Abella in a fluent and high level manner, where all the meta-theory can be proven in a straightforward manner. We illustrate its use in giving a clean formulation of fact that the Howe extension of applicative similarity is a pre-congruence, a standard result from the meta-theory of the λ-calculus that requires sophistication in treating simultaneous substitutions.

7.10. Hybrid Linear Logic, revisited

Participant: Kaustuv Chaudhuri.

Hybrid Linear Logic (HyLL) was proposed by Chaudhuri and Despeyroux in 2010 as a meta-logic for reasoning about constrained transition systems, with applications to a number of domains including formal molecular biology [36]. This logic is an extension of (intuitionistic) linear logic with hybrid connectives that can reason about monoidal constraint domains such as instants of time or rate functions. *Linear logic with subexponential* is a different extension of linear logic that has been proposed as a mechanism for capturing certain well known constrained settings such as bigraphs [39] or concurrent constraint programming [65]. In a paper accepted to MSCS [5] we show how to relate these two extensions of linear logic by giving an embedding of HyLL into linear logic with subexponentials. Furthermore, we show that subexponentials are able to give an adequate encoding of CTL$*$, which is beyond the expressive power of HyLL. Thus, subexponentials appear to be the better choice as a foundation for constraints in linear logic.

7.11. Proof Nets and the Linear Substitution Calculus

Participant: Beniamino Accattoli.

This work [21] belongs to line of work *Cost Models and Abstract Machines for Functional Programs*, supported by the ANR project COCA HOLA, and it has been published in the proceedings of the international conference ICTAC 2018.

The *Linear Substitution Calculus* (LSC) is a refinement of the λ-calculus that is crucial for the study of cost models for functional programs, as it enables a sharp and yet simple decomposition of the evaluation of λ-terms, and it is employed in the proof of various results about cost models in the literature.

In this work we show that the LSC is isomorphic to the linear logic representation of the λ-calculus. More precisely, it is isomorphic to the *proof nets* presentation of such a fragment of linear logic. Proof nets are a graphical formalism, which—as most graphical formalisms—is handy for intuitions but not prone to formal reasoning. The result is relevant because it allows to manipulate formally a graphical formalism (proof nets) by means of an ordinary term syntax (the LSC).

7.12. Tight Typings and Split Bounds

Participants: Beniamino Accattoli, Stéphane Graham-Lengrand.

This joint work with Delia Kesner (Paris Diderot University) [12] belongs to line of work *Cost Models and Abstract Machines for Functional Programs*, supported by the ANR project COCA HOLA, and it has been published in the proceedings of the international conference ICFP 2018.

Intersection types are a classic tool in the study of the λ-calculus. They are known to characterise various termination properties.

It is also well-known that *multi types*, a variant of intersection types strongly related to linear logic, also characterise termination properties. Typing derivation of multi types, moreover, provide quantitative information such as the number of evaluation step and the size of the results, as first shown by de Carvalho.

In this work we provide some new results on this line of work, notably we provide the first quantitative study via multi types of the leftmost and linear head evaluation strategies. Moreover, we show that our approach covers also the other cases in the literature.

7.13. Types of Fireballs

Participant: Beniamino Accattoli.

This joint work with Giulio Guerrieri (Bologna University) [22] belongs to line of work *Cost Models and Abstract Machines for Functional Programs*, supported by the ANR project COCA HOLA, and it has been published in the proceedings of the international conference APLAS 2018.

The theory of the call-by-value λ-calculus has mostly been developed for *closed* programs, that is, programs without free variables. In the last few years, the authors dedicated considerable efforts to extend it to open terms, that is the case relevant for the implementation of proof assistants. The simplest presentation of the call-by-value λ-calculus for open terms is the *fireball calculus*.

In this work we extend the quantitative study via multi types mentioned in *Tight Typings and Split Bounds* to the fireball calculus.

7.14. Decision procedures for intuitionistic propositional logic

Participant: Stéphane Graham-Lengrand.

Provability in intuitionistic propositional logic is decidable and, as revealed by the works of, e.g., Vorobev [72], Hudelmaier [51] and Dyckhoff [42], proof theory can provide natural decision procedures, which have been implemented in various software. More precisely, a decision procedure is obtained by performing direct root-first proof-search in (different variants of) a sequent calculus system called LJT (aka G4ip); termination is ensured by a property of the sequent calculus called depth-boundedness.

Independently from this, Claessen and Rosen [40] recently proposed a decision procedure for the same logic, based on a methodology used in the field of Satisfiability-Modulo-Theories (SMT). Their implementation clearly outperforms the sequent-calculus-based implementations.

In 2018 we managed to establish of formal connection between the G4ip sequent calculus and the algorithm from [40], revealing the features that they share and the features that distinguish them. This connection is interesting because it gives a proof-theoretical light on SMT-solving techniques, and it opens the door to the design of an intuitionistic version of the CDCL algorithm used in SAT-solvers, which decides provability in classical logic.

7.15. Admissible Tools in the Kitchen of Intuitionistic Logic

Participants: Matteo Manighetti, Andrea Condoluci.

In this work we study the computational meaning of the inference rules that are admissible, but not derivable, in intuitionistic logic [16].

An inference rule is admissible for a logic if whenever its antecedent is derivable, its conclusion was already derivable without the rule. In classical logic, whenever this is the case, then also the implication between antecedent and conclusion is derivable. The notion of an admissible rule is therefore internalized in the logic.

This is not the case for intuitionistic logic, and some rules that are admissible are not derivable: therefore they need reasoning outside the usual intuitionistic logic in order to be reduced to purely intuitionistic derivation.

In this work we propose a proof system with term annotations and reduction rules to give a computational meaning to these reductions.

PI.R2 Project-Team

6. New Results

6.1. Effects in proof theory and programming

Participants: Hugo Herbelin, Yann Régis-Gianas, Alexis Saurin, Exequiel Rivas Gadda.

6.1.1. Interfaces for computational effects

Exequiel Rivas studied the relation between interfaces for computational effects in programming languages: arrows, idioms and monads. Building on previous results of Lindley, Yallop and Wadler, a categorical account was developed by means of monoidal adjunctions. This work was presented in MSFP 2018 [40] and later in SYCO I. Together with Ruben Pieters and Tom Schrijvers, a journal version of the article is currently being prepared that includes this work and previous work on non-monadic handlers. It will be submitted to the Journal of Functional Programming.

6.1.2. Monads with merging

In collaboration with Mauro Jaskelioff, Exequiel Rivas developed monads with merge-like operators. These operators are based on two well-known algebraic theories for concurrency: classic process algebras and the more recent concurrent monoids. This resulted in an article submitted to FoSSaCS.

6.1.3. Relative effects: coherence for skew structures

In joint work with Mauro Jaskelioff, Tarmo Uustalu and Niccolò Veltri, Exequiel Rivas developed coherence theorems in the setting of categories with skew structures: skew monoidal categories, skew near-rig categories, skew semigroup categories. These skew structures are motivated by the study of relative effects in programming languages, where the primary example are relative monads. The results are formalised in the programming language Agda. A journal article is currently being written.

6.1.4. Effectful proving

Hugo Herbelin started a program of reconstruction of different levels of computational strength of logic by means of translation to a core logic of polarised linear connectives.

6.1.5. On the computational strength of choice axioms

With the goal of transferring the effectful computational contents of the dependent choice to other forms of choice or bar induction axioms, Hugo Herbelin worked at clarifying the folklore regarding the strengths of various forms of choice and of bar induction.

In collaboration with Boban Velickovic, Alexis Saurin advised the LMFI master internship of Ikram Cherigui on classical realisability and forcing in set theory.

6.1.6. Effectful systems in Coq

In collaboration with Thomas Letan (Agence Nationale pour la Sécurité des Systèmes Informatiques), Pierre Chifflier (ANSSI) and Guillaume Hiet (Centrale Supélec), Yann Régis-Gianas developed a new approach to model and verify effectful systems in Coq. This work has been presented at FM 2018 [38].

6.2. Reasoning and programming with infinite data

Participants: Yann Régis-Gianas, Alexis Saurin, Abhishek De, Luc Pellissier, Xavier Onfroy.

This theme is part of the ANR project Rapido (see the National Initiatives section) which goes until end of september 2019.

6.2.1. Proof theory of infinitary and circular proofs

In collaboration with David Baelde, Amina Doumane, Guilhem Jaber and Denis Kuperberg, Alexis Saurin extended the proof theory of infinite and circular proofs for fixed-point logics in various directions by relaxing the validity condition necessary to distinguish sound proofs from invalid ones. The original validity condition considered by Baelde, Doumane and Saurin in CSL 2016 rules out lots of proofs which are computationally and semantically sound and does not account for the cut-axiom interaction in sequent proofs.

In the setting of sequent calculus, Saurin introduced together with Baelde, Doumane and Jaber a relaxed 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. The most general form of this criterion does not ensure productivity due to a discrepancy between the sequential nature of proofs in sequent calculus and the parallel nature of threads. Several directions of research have therefore been investigated from that point:

- In sequent calculus, Baelde, Doumane and Saurin provided a slight restriction of the full bouncing validity which grants productivity and validity of the cut-elimination process. This restriction still strictly extends previous notions of validity and is actually expressive enough to be undecidable as proved together with Kuperberg. Decidability can be recovered by constraining the shapes of bounces. Doumane and Saurin were able in the fall 2018 to generalise the CSL proof technique to be applicable to bouncing threads. Those results are currently being written targetting a submission early 2019.
- In the setting of natural deduction, Saurin and Jaber introduced a validity criterion aiming at ensuring productivity of a circular λ-calculus with inductive and coinductive types.
- In the fall 2018, Abhishek De started his PhD under Saurin's supervision. The first part of his PhD work is dedicated to lifting the proof theory of circular and infinitary proofs to the setting of proof nets, in which the bouncing criterion will be much more convenient to work with since the discrepancy between sequent proofs and parallel threads will be dealt with.

6.2.2. 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. New results on this topics followed in 2018.

In a work with Nollet and Tasson, Saurin published in CSL 2018 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 [46]. 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.

6.2.3. Streams and classical logic

Luc Pellissier started a post-doc in december 2018 funded by the RAPIDO project and started working with Alexis Saurin on the stream interpretation of $\Lambda\mu$-calculi by investigating the connection between $\Lambda\mu$-calculus and the parsimonious λ-calculus.

6.2.4. Formalising circular proofs and their validity condition

During the spring and summer 2018, Saurin started with Xavier Onfroy a formalisation of circular proofs in Coq. Until now, Onfroy formalised parity-automata and their meta-theory as a first step to capture the decidability condition of circular proofs. Preliminary formalisations of circular proofs have been considered by Onfroy but shall still be pursued in order to fit into the picture.

6.3. Effective higher-dimensional algebra

Participants: Antoine Allioux, Pierre-Louis Curien, Eric Finster, Yves Guiraud, Cédric Ho Thanh, Matthieu Sozeau.

6.3.1. Rewriting methods in algebra

Yves Guiraud has written with Philippe Malbos (Univ. Lyon 1) a survey on the use of rewriting methods in algebra, centered on a formulation of Squier's homotopical and homological theorems in the modern language of higher-dimensional categories. This article is intended as an introduction to the domain, mainly for graduate students, and has appeared in Mathematical Structures in Computer Science [32].

Yves Guiraud has completed a four-year collaboration with Eric Hoffbeck (Univ. Paris 13) and Philippe Malbos (Univ. Lyon 1), 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 [71] 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 [70] and Poincaré-Birkhoff-Witt bases [107]. Then, they transposed the construction of [14], based on an extension of Squier's theorem [110] 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 will appear in Mathematische Zeitschrift [31].

Yves Guiraud has written his "Habilitation à diriger des recherches" manuscript, as a survey on rewriting methods in algebra based on Squier theory [13]. The defense is planned for Spring 2019.

Yves Guiraud works with Dimitri Ara (Univ. Aix-Marseille), Albert Burroni, Philippe Malbos (Univ. Lyon 1), François Métayer (Univ. Nanterre) and Samuel Mimram (École Polytechnique) on a reference book on the theory of polygraphs and higher-dimensional categories, and their applications in rewriting theory and homotopical algebra.

Yves Guiraud works with Marcelo Fiore (Univ. Cambridge) on the theoretical foundations of higher-dimensional algebra, in order to develop a common setting to develop rewriting methods for various algebraic structures at the same time. Practically, they aim at a definition of polygraphic resolutions of monoids in monoidal categories, based on the recent notion of n-oid in an n-oidal category. This theory will subsume the known cases of monoids and associative algebras, and encompass a wide range of objects, such as Lawvere theories (for term rewriting), operads (for Gröbner bases) or higher-order theories (for the λ-calculus).

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, Cédric Ho Thanh and Samuel Mimram have developped (in several variants) a type-theoretical treatment of opetopes and finite opetopic sets, and have shown that the models of their type theory are indeed the opetopic sets as defined mathematically by the above authors. This work is being submitted to an international conference. Also, Cédric Ho Thanh has given a direct precise proof of the equivalence between many-to-one polygraphs and opetopic sets, thus establishing a connection with the theory of polygraphs [57].

6.3.2. Garside methods in algebra and rewriting

Building on [9], Yves Guiraud is currently finishing with Matthieu Picantin (Univ. Paris 7) a work that generalises already known constructions such as the bar resolution, several resolutions defined by Dehornoy and Lafont [79], 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.

6.3.3. Foundations and formalisation of higher algebra

Antoine Allioux (PhD started in February), Eric Finster, Yves Guiraud and Matthieu Sozeau are exploring the development of higher algebra in type theory. To formalise higher algebra, one needs a new source of coherent structure in type theory. Finster has developed an internalisation of polynomial monads (of which opetopes and ∞-categories are instances) in type theory, which ought to provide such a coherent algebraic structure, inspired by the work of Kock et al [96]. Antoine Allioux is focusing on building an equivalence of types between categories seen as polynomial monads and the standard univalent categories in Homotopy Type Theory [22]. Another result that should follow is the ability to define simplicial types in Homotopy Type Theory, a long standing open problem in the field. An article on this subject is in preparation. Once armed with such a definition mechanism for higher algebraic structures and their algebras, it should be possible to internalise results from higher rewriting theory in type theory, which was the initial goal of this project.

6.3.4. Type Theory and Higher Topos Theory

Eric Finster explored the connections between intensional type theory and the theory of higher topoi, as developed in the works on Joyal and Lurie [103]. In particular, in collaboration with Mathieu Anel, André Joyal and Georg Biedermann, he gave a proof of a new result about the generation of left exact modalities in higher topoi, which has a corresponding internalisation in Homotopy Type Theory. Applications of this result to the Goodwillie Calculus, an advanced technique in abstract homotopy theory, resulted in the article [28].

6.4. Incrementality

Participants: Thibaut Girka, Yann Régis-Gianas.

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 has been submitted to ESOP 2019.

In collaboration with Olivier Martinot (Paris Diderot), Yann Régis-Gianas studied a new technique to implement incrementalised operations on lists. A paper is to be submitted to ICFP 2019.

6.4.2. Difference languages

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, Gaëtan Gilbert, Cyprien Mangin, Théo Winterhalter, Théo Zimmermann, Thierry Martinez.

6.5.1. Homotopy type theory

Hugo Herbelin developed the syntax for a variant of Cohen, Coquand, Huber and Mörtberg's Cubical Type Theory where equality on types is defined to be equivalence of types, thus satisfying univalence by construction.

6.5.2. Proof irrelevance and Homotopy Type Theory

Gaëtan Gilbert (PhD student of N. Tabareau, Gallinette and M. Sozeau) continued developing the theory and implementation of *strict* propositions in the calculus of inductive constructions. In collaboration with Jesper Cockx (Chalmers), they developed this notion in full in an article at POPL 19 [30]. Strict propositions enjoy definitional proof-irrelevance and are compatible with both Univalence and Uniqueness of Identity Proofs, providing a foundation for further research in both directions: dealing with strict structures in homotopy type theory, and improving the support for programming with dependent types and proofs. They have shown in particular how to translate inductive types that can be seen as strict propositions into recursively defined types, providing a fix to the "singleton elimination" criterion used in Coq to treat the interaction of propositions (in Prop) and informative objects (in Type). Together with Pierre Letouzey, Matthieu Sozeau is pursuing an adaptation of the Prop sort informed by this new result. In particular, Pierre Letouzey is now experimenting with alternative ways to handle the accessibility arguments of Coq general fixpoints during extraction. Historically, the elimination of these arguments was a consequence of the accessibility inductive type being in Prop. But this can actually be seen as a more general dead-code elimination method. This leverages the need for accessibility to be in sort Prop, and hence opens new prospects concerning the Prop universe and the proof irrelevance.

6.5.3. Extensionality and Intensionality in Type Theory

Théo Winterhalter, Nicolas Tabareau and Matthieu Sozeau studied and formalised a complete translation from Extensional to Intensional Type Theory in Coq, now published at CPP 2019 [43]. They show that, contrary to the original paper proof of Oury, the target intensional type theory only needs to be extended with the Uniqueness of Identity Proofs principle and Functional Extensionality, settling concretely and formally a question that was studied semantically and up-to now only on paper by Hofmann and Altenkirch [61]. The translation was formalised using the Template-Coq framework and gives rise to an executable translation from partial terms of ETT into terms of Coq annotated with transports of equalities. This provides a simple way to justify the consistency of type theories extending the definitional equality relation by provable propositional equalities, and shows the equivalence of 2-level type theory [62] and the Homotopy Type System proposed by Voevodsky.

6.5.4. Dependent pattern-matching and recursion

Cyprien Mangin and Matthieu Sozeau have continued work on the Equations plugin of Coq, Equations now provides means to define nested, mutual and well-founded recursive definitions, together with a definitional compilation of dependent-pattern matching avoiding the use of axioms. In recent work, Matthieu Sozeau uncovered a new way to deal with dependent pattern-matching on inductive families avoiding more uses of the K axiom, inspired by the work of Cockx [74], that integrates well with the simplification engine developed for Equations. An article describing this work is in revision [58].

Thierry Martinez continued 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. Explicit Cumulativity

Pierre Letouzey continued exploring with the help of Matthieu Sozeau a version of Coq's logic (CIC) where the cumulativity rule is explicit. This cumulativity rule is a form of coercion between Coq universes, and is done silently in Coq up to now. Having a version of CIC where the use of the cumulativity bewteen Prop and Type is traceable would be of great interest. In particular this would lead to a solid ground for the Coq extraction tool and solve some of its current limitations. Moreover, an explicit cumulativity would also help significantly the studies of Coq theoretical models. A prototype version of Coq is now available, but only a fragment of the standard library has been adapted to explicit cumulativity. In particular, the equalities of equalities currently need some amending, and this process is quite cumbersome.

6.5.6. Cumulativity for Inductive Types

Together with Amin Timany, Matthieu Sozeau developed the Calculus of Cumulative Inductive Constructions which extends the cumulativity relation of universes to universe polymorphic inductive types. This work was presented at FSCD 2018 [42]. The development of the model of this calculus suggested a refinement of the implementation which was integrated in Coq 8.8, providing a more flexible subtyping relation on inductive types in Coq. Notably, this work shrinks the gap to emulate the so-called "template" polymorphism of Coq with cumulative universe polymorphism. Cumulative Inductive Types also provide an apropriate basis to formalise the notions of small and large categories in type theory, avoiding the introduction of coercions. In particular, it provides a way to define a well-behaved category of types and functions and constructions on it, like the Yoneda embedding, which would not be expressible without cumulativity. Finally, Cumulative Inductive Types allow the definition of syntactic models of type theories with cumulativity inside Coq, as pioneered by Boulier *et al* [69].

6.5.7. Mathematical notations in Coq

Hugo Herbelin developed new extensions of the system of mathematical notation of Coq: support for autonomous auxiliary grammars, support for binders over arbitrary patterns, support for generic notations for applications.

6.5.8. Software engineering aspects of the development of Coq

Théo Zimmermann has studied software engineering and open collaboration aspects of the development of Coq.

Following the migration of the Coq bug tracker from Bugzilla to GitHub which he conducted in 2017, he analyzed data (extracted through the GitHub API), in collaboration with Annalí Casanueva Artís from the Paris School of Economics. The results show an increased number of bugs by core developers and an increased diversity of the people commenting bug reports. These results validate *a posteriori* the usefulness of such a switch. A paper [60] has been written and has been presented at the EAQSE workshop (without proceedings). The current objective is to publish the paper in the MSR 2019 conference.

Following discussions dating back from the end of 2017, he has founded the coq-community GitHub organisation in July 2018. This is a project for a collaborative, community-driven effort for the long-term maintenance and advertisement of Coq packages. Already 10 pre-existing Coq projects (plugins and libraries) have been moved to this organisation since then (seven of them are former Coq contribs that were fixed from time to time by the Coq developers themselves – mostly by Hugo Herbelin). The organisation also hosts a "manifesto" repository for general discussion, documentation and advice to developers (including already a few reusable templates for Coq projects), and a docker-coq project to provide reusable Docker images with Coq. The next objectives are to get started on the collaborative documentation (starting with a work by Pierre Castéran from LaBRI) and to create an editorial committee. Théo Zimmermann and Yann Régis-Gianas are preparing an article of the model proposed by the various existing *-community GitHub organisations (including the elm-community organisation from which coq-community was inspired, and ocaml-community which was influenced by coq-community itself).

In addition, Théo Zimmermann has coordinated efforts to improve the documentation of Coq, has documented the release process that he had put in place with Maxime Dénès, and has developed a GitHub / GitLab bot (in OCaml) that is used to automatise many useful functions for the Coq development (continuous integration and backporting of pull requests in particular). The goal is to make this bot modular and reusable for other projects.

6.5.9. Coordination of the development of Coq

The amount of contributions to the Coq system increased significantly in the recent years (around 50 pull-requests are reviewed, discussed and merged each month, approximately). Hugo Herbelin, Matthieu Sozeau and Théo Zimmermann, helped by members from Gallinette (Nantes) and Marelle (Sophia-Antipolis), devoted an important part of their time to coordinate the development, to review propositions of extensions of Coq from external and/or young contributors, and to propose themselves extensions (see the corresponding paragraphs).

6.6. Formalisation and verification

Participants: Pierre-Louis Curien, Kailiang Ji, Pierre Letouzey, Jean-Jacques Lévy, Cyprien Mangin, Daniel de Rauglaudre, Matthieu Sozeau.

6.6.1. Proofs and surfaces

Following ideas of J. Richter-Gebert, Pierre-Louis Curien, together with Jovana Obradović (former PhD student of the team and now postdoc in Prague), joined a project with Zoran Petrić and other Serbian colleagues on formalising proofs of incidence theorems (arising by repeated use of Menelaus theorem) by means of a cyclic sequent calculus, by which is meant that a (proof of a) sequent $\vdash \Gamma$ stands for the conjunction of all (proofs of) traditional sequents $\Gamma \setminus \psi \vdash \psi$. We have designed a proof system, showed its soundness, and experimented it on an extended set of examples from elementary projective geometry. A paper is being written.

6.6.2. Hofstadter nested recursive functions and Coq

Pierre Letouzey continued this year the study of a family of nested recursive functions proposed by D. Hofstadter in his book "Gödel Escher Bach". This is a generalisation of the earlier work [20], bringing a large number of new insights as well as many new conjectures. Most of the work is already certified in Coq, with generalised and/or nicer proofs, see https://www.irif.fr/~letouzey/hofstadter_g/. Many interactions with Fibonacci numbers or similar recursive sequence have been found. Pierre Letouzey even stumbled upon a Rauzy fractal during this investigation, which is still ongoing.

6.6.3. Real Numbers in Coq

The present Coq library of real numbers is made of 17 axioms. Daniel de Rauglaudre has been studying the possibility of making an implementation with one only axiom: the Limited Principle of Omniscience (LPO) which says that we can differentiate an infinite sequence of 0s from an infinite sequence holding something else than 0 (it seems obvious but it cannot be proved in constructive logic). This axiom had been already used in the formal proof of Puiseux' theorem done some years ago (only axiom of this proof too).

Real numbers are defined by an infinite sequence of digits and the operations of addition and multiplication by algorithms using LPO.

It was tested in OCaml, the axiom being replaced by a function having a limit corresponding to the precision of the computation and it seems to work. But the proof in Coq that this implementation is a field stumbles on difficulties about the associativity of addition which is more complicated than expected. Several tracks have been experimented with Hugo Herbelin's help.

6.6.4. Proofs of algorithms on graphs

Jean-Jacques Lévy and Chen Ran (a PhD student at the Institute of Software, Beijing) pursue their work about formal proofs of graph algorithms. Their goal is to provide proofs of algorithms checked by computer and 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 (Kosaraju - 1978 and Tarjan - 1972). Their proofs use the multi-sorted first-order logic with inductive predicates of the Why3 system (research-team Toccata, Saclay). They also widely use the numerous automatic provers interfaced with Why3. A very minor part of these proofs is also achieved in Coq. The difficulty of this approach is to combine automatic provers and the intuitive design. Another point is to define the good level of abstraction in order to avoid too many implementation features while keeping an effective presentation.

In 2017, the same proofs were fully completed in Coq-ssreflect with the Mathematical Components library by Cohen and Théry (research-team Marelle, Sophia-Antipolis), and in Isabelle-HOL by Merz (research-team VeriDis, Nancy), both proofs with the assistance of J.-J. Lévy. These proofs are between a factor 3 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. On the way, this collaboration led to a new, better presentation of the Why3 proof.

Part of this work (Tarjan 1972) was presented at JFLA 2017, a more comprehensive version was presented at the VSTTE 2017 conference in Heidelberg. Scripts of proofs can be found at http://jeanjacqueslevy.net/why3, where other proofs of graph algorithms are also present: acyclicity test, articulation points, biconnected components. A proof of Tarjan's planarity test is also under design. A paper entitled “Formal Proofs of Tarjan's Algorithm in Why3, Coq and Isabelle” is under submission to a conference.

6.6.5. 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 from Coq's Gallina language down to CompCert C-light which provides itself a certified compilation path to assembly language. Matthieu Sozeau focused on the front-end part of CertiCoq, providing formal proofs of the first two phases of the compiler. The first phase translates from Coq syntax to a more amenable representation for metatheoretical study, and the second phase performs extraction to an untyped lambda-calculus with datatypes and mutual (co-)fixpoints. These two phases are of general use and are now integrated and developed in the MetaCoq project. The CertiCoq team expects to release a first version of the compiler in the beginning of 2019, along with an article describing it.

MetaCoq is a project led by Matthieu Sozeau, in collaboration with Simon Boulier and Nicolas Tabareau in Nantes, Abhishek Anand and Gregory Malecha (BedRock Systems, Inc) and Yannick Forster in Saarbrucken. The project was born from the extension of the Template-Coq reification plugin of G. Malecha, which now contains:

- A specification of the typing rules of Coq and its basic metatheoretical properties (weakening, substitution). This specification is not entirely complete yet, as the positivity and guard-checking of definitions is missing. Cyprien Mangin has formalised the regular tree structure used by the guard checker, and a simple positivity check for inductive types. Its integration is ongoing.
- A (partial) proof of the correctness and completeness of a reference type-checker with respect to these rules.
- An implementation of the extraction phase of Coq, which is used in the CertiCoq project. The proof of "syntactic" correctness of this phase, that is the preservation of weak call-by-value reduction by extraction is ongoing.
- A monad giving the ability to program arbitrary plugins in Coq itself, in the style of MTac.

. The foundation of this project was published at ITP 2018 [37], and a journal article is in preparation.

In collaboration with Jan-Oliver Kaiser (MPI-SWS), Beta Ziliani (CONICET/FAMAF), Robbert Krebbers (ICIS) and Derek Dreyer (MPI-SWS), Yann Régis-Gianas participates in the Mtac2 project, a metaprogramming language for Coq. The new version of this language has been presented at ICFP 2018 [34]. It includes in particular in a depedently-typed variant of the LCF tactic typing discipline.

In collaboration with Xavier Denis (Paris Diderot), Yann Régis-Gianas is implementing a compiler for Mtac2.

6.6.6. Equivalences for free!

Nicolas Tabareau (Inria Nantes), Eric Tanter (U. Chile in Santiago) and Matthieu Sozeau developed a new parametricity translation for justifying the transport of programs and proofs by equivalences in type theory [36]. Inspired by the Univalence axiom, they show that every construction of type theory (minus inductive families indexed by universes) respect type equivalence, and provide a modified parametricity translation that can be used to construct the proof of invariance by equivalence of any term. This translation is engineered so that transports do not appear during this inference, allowing an easy implementation of a transfer metaprogram in type theory using type class inference. Using this metaprogram, one can automatically transport libraries of implementations and their proofs from one type to an equivalent one, including cases where dependent types are used. While the translation ultimately relies on the univalence axiom to treat universes, its use can be avoided in many cases, providing an effective translation that can be evaluated inside type theory.

6.6.7. Detecting K-Synchronisability Violations

Ahmed Bouajjani, Constantin Enea, Kailiang Ji and Shaz Qadeer introduced a bounded analysis that explores a special type of computations, called k-synchronous, for analyzing message passing programs. They gave a procedure for deciding *k-synchronisability* of a program, i.e., whether every computation is equivalent (has the same happens-before relation) to one of its k-synchronous computations. They also showed that reachability over k-synchronous computations and checking k-synchronisability are both PSPACE-complete. Furthermore, they introduced a class of programs called *flow-bounded* for which the problem of deciding whether there exists a $k > 0$ for which the program is k-synchronisable, is decidable. The *k-synchronisability* violation detection algorithm was implemented in Spin model checker. This work was published at CAV 2018 [48].

SUMO Project-Team

7. New Results

7.1. Analysis and Verification of Quantitative Systems

7.1.1. Verification of Concurrent Timed Systems

Participants : Éric Fabre, Loïc Hélouët, Karim Kecir

7.1.1.1. Combining Free Choice and Time in Petri Nets

Time Petri nets (TPNs) are a classical extension of Petri nets with timing constraints attached to transitions, for which most verification problems are undecidable. In [3], We consider TPNs under a strong semantics with multiple enablings of transitions. We focus on a structural subclass of unbounded TPNs, where the underlying untimed net is free choice, and show that it enjoys nice properties in the timed setting under a multi-enabling semantics. In particular, we show 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 consider 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-enabling semantics, we show decidability of robustness of firability and of termination under both guard enlargement and shrinking.

7.1.1.2. Production Systems with Concurrent Tasks

The work in [7] considers the realizability of expected schedules by production systems with concurrent tasks, bounded resources that have to be shared among tasks, and random behaviors and durations. Schedules are high level views of desired executions of systems represented as partial orders decorated with timing constraints. Production systems (production cells, train networks...) are modeled as stochastic time Petri nets STPNs with an elementary (1-bounded) semantics. We first propose a notion of time processes to give a partial order semantics to STPNs. We then consider 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 hence consider probabilistic realizability up to a time units, that holds if the probability that N realizes S with constraints enlarged by a 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.

7.1.2. Testing of Timed Systems

Participants : Léo Henry, Thierry Jéron, Nicolas Markey

Partial observability and controllability are two well-known issues in test-case synthesis for interactive systems. In [25], we address the problem of partial control in the synthesis of test cases from timed-automata specifications. Building on the tioco timed testing framework, we extend a previous game interpretation of the test-synthesis problem from the untimed to the timed setting. This extension requires a deep reworking of the models, game interpretation and test-synthesis algorithms. We exhibit strategies of a game that tries to minimize both control losses and distance to the satisfaction of a test purpose, and prove they are winning under some fairness assumptions. This entails that when turning those strategies into test cases, we get properties such as soundness and exhaustiveness of the test synthesis method.

7.1.3. Analysis of Stochastic Systems

Participants : Nathalie Bertrand

A decade ago, Abdulla, Ben Henda and Mayr introduced the elegant concept of decisiveness for denumerable Markov chains. Roughly speaking, decisiveness allows one to lift most good properties from finite Markov chains to denumerable ones, and therefore to adapt existing verification algorithms to infinite-state models. Decisive Markov chains however do not encompass stochastic real-time systems, and general stochastic transition systems (STSs for short) are needed. In [4], we provide a framework to perform both the qualitative and the quantitative analysis of STSs. First, we define various notions of decisiveness, notions of fairness and of attractors for STSs, and make explicit the relationships between them. Then, we define a notion of abstraction, together with natural concepts of soundness and completeness, and we give general transfer properties, which will be central to several verification algorithms on STSs. We further design a generic construction which will be useful for the analysis of ω-regular properties, when a finite attractor exists, either in the system (if it is denumerable), or in a sound denumerable abstraction of the system. We next provide algorithms for qualitative model-checking, and generic approximation procedures for quantitative model-checking. Finally, we instantiate our framework with stochastic timed automata (STA), generalized semi-Markov processes (GSMPs) and stochastic time Petri nets (STPNs), three models combining dense-time and probabilities. This allows us to derive decidability and approximability results for the verification of these models. Some of these results were known from the literature, but our generic approach permits to view them in a unified framework, and to obtain them with less effort. We also derive interesting new approximability results for STA, GSMPs and STPNs.

7.1.4. Opacity for Quantitative Systems

Participants : Loïc Hélouët, Hervé Marchand

7.1.4.1. Quantitative Opacity

The work in [26] considers quantitative approaches for opacity. A system satisfies opacity if its secret behaviors cannot be detected by any user of the system. Opacity of distributed systems was originally set as a boolean predicate before being quantified as measures in a probabilistic setting. This paper considers a different quantitative approach that measures the efforts that a malicious user has to make to detect a secret. This effort is measured as a distance w.r.t a regular profile specifying a normal behavior. This leads to several notions of quantitative opacity. When attackers are passive that is, when they just observe the system, quantitative opacity is brought back to a language inclusion problem, and is PSPACE-complete. When attackers are active, that is, interact with the system in order to detect secret behaviors within a finite depth observation, quantitative opacity turns out to be a two-player finite-state quantitative game of partial observation. A winning strategy for an attacker is a sequence of interactions with the system leading to a secret detection without exceeding some profile deviation measure threshold. In this active setting, the complexity of opacity is EXPTIME-complete.

7.1.4.2. Opacity with Powerful Attackers

In [27], we consider state-based opacity in a setting where attackers of a secret have additional observation capabilities allowing them to know which inputs are allowed by a system. This capability allows attackers of a system to partially disambiguate the possible set of states the system might be in, and increases the power of an attacker. We show that regular opacity (opacity of a property described by a regular language) is decidable in this setting. We then address the question of controlling a system so that it becomes opaque, and solve this question by recasting the problem in a game setting.

7.1.5. Diagnosis of Quantitative Systems

Participants : Blaise Genest, Éric Fabre, Hugo Bazille, Nicolas Markey

7.1.5.1. Diagnosis for Timed Automata

In [20], we consider the problems of efficiently diagnosing and predicting what did (or will) happen in a partially-observable one-clock timed automaton. We introduce timed sets as a formalism to keep track of the evolution of the reachable configurations over time, and build a candidate diagnoser for our timed automaton. We report on our implementation of this approach compared to the algorithm of Tripakis, *Fault diagnosis for timed automata*, 2002.

7.1.5.2. *Quantitative Diagnosis for 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 in [17]. We are interested in quantifying how fast the diagnosability can be performed. For that, we give an algorithm to compute in polynomial time any moment of the distribution of the detection delay. This allows one to approximate the distribution of detection delay, and to provide lower bounds on the probability that detection takes place at most T events after the fault.

One problem with A-diagnosability is that in the worst case, a subset construction needs to be performed, leading to an exponential blow-up in the number of states. To mitigate this, we proposed in [16] different techniques that avoid this blow-up in a large number of cases.

7.2. Control of Quantitative Systems

7.2.1. Reactive Synthesis for Quantitative Systems

Participants : Hervé Marchand, Nicolas Markey

7.2.1.1. *Optimal and Robust Controller Synthesis*

We propose a novel framework for the synthesis of robust and optimal energy-aware controllers. The framework is based on energy timed automata, allowing for easy expression of timing-constraints and variable energy-rates. We prove decidability of the energy-constrained infinite-run problem in settings with both certainty and uncertainty of the energy-rates. We also consider the optimization problem of identifying the minimal upper bound that will permit existence of energy-constrained infinite runs. Our algorithms are based on quantifier elimination for linear real arithmetic. Using Mathematica and Mjollnir, we illustrate our framework through a real industrial example of a hydraulic oil pump. Compared with previous approaches our method is completely automated and provides improved results.

7.2.1.2. *Average-Energy Games*

Two-player quantitative zero-sum games provide a natural framework to synthesize controllers with performance guarantees for reactive systems within an uncontrollable environment. Classical settings include mean-payoff games, where the objective is to optimize the long-run average gain per action, and energy games, where the system has to avoid running out of energy. In [5], we study average-energy games, where the goal is to optimize the long-run average of the accumulated energy. We show that this objective arises naturally in several applications, and that it yields interesting connections with previous concepts in the literature. We prove that deciding the winner in such games is in NP∩coNP and at least as hard as solving mean-payoff games, and we establish that memoryless strategies suffice to win. We also consider the case where the system has to minimize the average-energy while maintaining the accumulated energy within predefined bounds at all times: this corresponds to operating with a finite-capacity storage for energy. We give results for one-player and two-player games, and establish complexity bounds and memory requirements.

7.2.1.3. *Compositional Controller Synthesis*

In [8], we present a correct-by-design method of state-dependent control synthesis for sampled switching systems. Given a target region R of the state space, our method builds a capture set S and a control that steers any element of S into R. The method works by iterated backward reachability from R. It is also used to synthesize a recurrence control that makes any state of R return to R infinitely often. We explain how the synthesis method can be performed in a compositional manner, and apply it to the synthesis of a compositional control for a concrete floor-heating system with 11 rooms and up to $2^{11} = 2048$ switching modes.

7.2.1.4. Symbolic Algorithms for Control

In [18], we put forward a new modeling technique for Dynamic Resource Management (DRM) based on discrete events control for symbolic logico-numerical systems, especially Discrete Controller Synthesis (DCS). The resulting models involve state and input variables defined on an infinite domain (Integers), thereby no exact DCS algorithm exists for safety control. We thus formally define the notion of limited lookahead, and associated best-effort control objectives targeting safety and optimization on a sliding window for a number of steps ahead. We give symbolic algorithms, illustrate our approach on an example model for DRM, and report on performance results based on an implementation in our tool ReaX.

7.2.2. Control of Stochastic Systems

Participants : Nathalie Bertrand, Blaise Genest, Nicolas Markey, Ocan Sankur

7.2.2.1. Multi-Weighted Markov Decision Processes

In [19], we study the synthesis of schedulers in double-weighted Markov decision processes, which satisfy both a percentile constraint over a weighted reachability condition, and a quantitative constraint on the expected value of a random variable defined using a weighted reachability condition. This problem is inspired by the modelization of an electric-vehicle charging problem. We study the cartography of the problem, when one parameter varies, and show how a partial cartography can be obtained via two sequences of opimization problems. We discuss completeness and feasability of the method.

7.2.2.2. Stochastic Shortest Paths and Weight-Bounded Reachability

The work in [14] deals with finite-state Markov decision processes (MDPs) with integer weights assigned to each state-action pair. New algorithms are presented to classify end components according to their limiting behavior with respect to the accumulated weights. These algorithms are used to provide solutions for two types of fundamental problems for integer-weighted MDPs. First, a polynomial-time algorithm for the classical stochastic shortest path problem is presented, generalizing known results for special classes of weighted MDPs. Second, qualitative probability constraints for weight-bounded (repeated) reachability conditions are addressed. Among others, it is shown that the problem to decide whether a disjunction of weight-bounded reachability conditions holds almost surely under some scheduler belongs to NP∩coNP, is solvable in pseudo-polynomial time and is at least as hard as solving two-player mean-payoff games, while the corresponding problem for universal quantification over schedulers is solvable in polynomial time.

7.2.2.3. Distribution-based Objectives for Markov Decision Processes

In the scope of associated team EQuaVE, we have considered quantitative control of stochastic systems [10]. More precisely, the aim is to control the MDP so that the distribution over states stays inside a safe polytope. This represents a trade off between perfect information (the system is in exactly one state) and no information (we need to consider the belief distribution over states, and further the action played by the controller cannot be based on the state). Interestingly, we get an efficient polynomial time complexity to check whether there exists a distribution from which there exists a controller keeping the MDP in the safe polytope. This is surprising as the same question from a given distribution is not known to be decidable, even if the controller is fixed. Also, we have a co-NP complexity for deciding whether for every initial distribution, there is controller keeping the distribution in the safe polytope. Finally, we showed that an alternate representation of the input polytope allows us to get a polynomial time algorithm for safety from all initial distributions.

7.3. Management of Large Distributed Systems

7.3.1. Parameterized Systems

Participants : Nathalie Bertrand, Nicolas Markey

Reconfigurable broadcast networks provide a convenient formalism for modelling and reasoning about networks of mobile agents broadcasting messages to other agents following some (evolving) communication topology. The parameterized verification of such models aims at checking whether a given property holds irrespective of the initial configuration (number of agents, initial states and initial communication topology). In [15], we focus on the synchronization property, asking whether all agents converge to a set of target states after some execution. This problem is known to be decidable in polynomial time when no constraints are imposed on the evolution of the communication topology (while it is undecidable for static broadcast networks).

During the internship of A.R. Balasubramanian, we investigated how various constraints on reconfigurations affect the decidability and complexity of the synchronization problem. In particular, we show that when bounding the number of reconfigured links between two communications steps by a constant, synchronization becomes undecidable; on the other hand, synchronization remains decidable in PTIME when the bound grows with the number of agents.

7.3.2. *Smart Regulation for Urban Trains*

Participants : Loïc Hélouët, Karim Kecir, Flavia Palmieri

We have launched a new thread of research for efficient regulation with the M2 internship of Flavia Palmieri. The objective is to use efficient planning techniques to perform regulation in metro networks. Usually, regulation algorithms are simple reactive rules, that build decisions from local measures of train delays. These algorithms are arbitrary decisions, which efficiency is only empirically proved. On the other hand, optimality of regulation decision with respect to some quality criterion could be achieved through optimization algorithms, associating an optimal execution date to next events (arrivals and departures) while fulfilling constraints on causal dependencies, track allocations, etc. However, these algorithms are NP-complete, and do not return answers fast enough to be used online as regulation tools (use usually expects a decision within a few seconds after a train's arrival). During this internship, we have started integrating optimal planning techniques to regulation schemes. The main idea is to perform optimization online for a subset of the next occurring events. Performance of this regulation scheme is currently under evaluation.

7.3.3. *Analysis of Concurrent Systems*

Participants : Éric Fabre, Loïc Hélouët, Engel Lefaucheux

7.3.3.1. Generalization of Unfolding Techniques for Petri Nets

The verification of concurrent systems relies on an adequate representation of their trajectory sets, where each trajectory is a partial order of events. Several compact structures have been proposed in the past, starting with unfoldings and event structures. While unfoldings expand both time and conflicts, they generate extremely large branching constructions. To avoid expanding conflicts where they are not meaningful, more compact structures were proposed, as merged processes and trellis processes. In [23], we examine structures that would not fully unfold time as well, thus resulting in partially unfolded nets. To do so, we proposed the notion of spread nets, (safe) Petri nets equipped with vector clocks on places and with ticking functions on transitions, and such that vector clocks are consistent with the ticking of transitions. Such nets allow one to generalize previous constructions as unfoldings and merged processes, and can be fully paremeterized to display or hide some behaviors of the net, and thus facilitate its analysis.

7.3.3.2. Hyper Partial Order Logic

In [21], we define HyPOL, a local hyper logic for partial order models, expressing properties of sets of runs. These properties depict shapes of causal dependencies in sets of partially ordered executions, with similarity relations defined as isomorphisms of past observations. This type of logics is tailored to address security properties of concurrent systems. Unsurprisingly, since comparison of projections are included, satisfiability of this logic is undecidable. We then address model checking of HyPOL and show that, already for safe Petri nets, the problem is undecidable. Fortunately, sensible restrictions of observations and nets allow us to bring back model checking of HyPOL to a decidable problem, namely model checking of MSO on graphs of bounded treewidth.

7.3.3.3. Diagnosability Analysis for Concurrent Systems

Petri nets have been proposed as a fundamental model for discrete-event systems in a wide variety of applications and have been an asset to reduce the computational complexity involved in solving a series of problems, such as control, state estimation, fault diagnosis, etc. Many of those problems require an analysis of the reachability graph of the Petri net. The basis reachability graph is a condensed version of the reachability graph that was introduced to efficiently solve problems linked to partial observation. It was in particular used for diagnosis which consists in deciding whether some fault events occurred or not in the system, given partial observations on the run of the system. However this method is, with very specific exceptions, limited to bounded Petri nets. In [28], we introduce the notion of basis coverability graph to remove this requirement. We then establish the relationship between the coverability graph and the basis coverability graph. Finally, we focus on the diagnosability problem: we show how the basis coverability graph can be used to get an efficient algorithm.

7.4. Data Driven Systems

7.4.1. Modular composition of Guarded Attribute Grammars

Participants : Éric Badouel

We investigate how the role of a user in a distributed collaborative systems modelled by a Guarded Attribute Grammar can be associated with a domain specific language (DSL) encapsulating a specific domain knowledge (expertise) and defining a set of services (a language-oriented approach). These DSLs communicate through service calls (a service-oriented approach).

Language oriented programming is an approach to software composition based on domain specific languages (DSL) dedicated to specific aspects of an application domain. In order to combine such languages we embed them into a host language (namely Haskell, a strongly typed higher-order lazy functional language). A DSL is then given by an algebraic type, whose operators are the constructors of abstract syntax trees. Such a multi-sorted signature is associated to a polynomial functor. An algebra for this functor tells us how to interpret the programs. Using Bekić's Theorem we defined in [13] a modular decomposition of algebras that leads to a class of parametric multi-sorted signatures, associated with regular functors, allowing for the modular design of DSLs.

In [12] we have addressed the problem of component reuse in the context of service-oriented programming and more specifically for the design of user-centric distributed collaborative systems modelled by Guarded Attribute Grammars. Following the contract-based specification of components we developp an approach to an interface theory for the roles in a collaborative system in three stages: we define a composition of interfaces that specifies how the component behaves with respect to its environement, we introduce an implementation order on interfaces and finally a residual operation on interfaces characterizing the systems that, when composed with a given component, can complement it in order to realize a global specification.

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 [91] [92]. 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 [23].

A Toolchain to Produce Correct-by-Construction OCaml Programs In the context of the research project Vocal, J.-C. Filliâtre, A. Paskevich, and M. Pereira, together with L. Gondelman (postdoc in January 2017) and S. Melo de Sousa (visiting Associate Professor from UBI, Portugal, in Sep/Oct 2017), designed and implemented a toolchain for the verification of OCaml code using Why3 [33]. In this framework, the user provides a formal specification within comments embedded in the OCaml interface file together with an implementation in Why3. Two tools automatically translate the former to a Why3 specification and the latter to an OCaml code. One the refinement proof is completed on the Why3 side, the overall diagram commutes, ensuring the soundness of the OCaml code.

Ghost monitors M. Clochard, C. Marché, and A. Paskevich designed a new approach to deductive program verification based on auxiliary programs called *ghost monitors*. This technique is useful when the syntactic structure of the target program is not well suited for verification, for example, when an essentially recursive algorithm is implemented in an iterative fashion. The approach consists in implementing, specifying, and verifying an auxiliary program that monitors the execution of the target program, in such a way that the correctness of the monitor entails the correctness of the target. This technique is also applicable when one wants to establish relational properties between two target programs written in different languages and having different syntactic structure [32] [29].

This approach is based on an earlier variant proposed in M. Clochard's PhD thesis [11]. The ghost monitor maintains the necessary data and invariants to facilitate the proof, it can be implemented and verified in any suitable framework, which does not have to be related to the language of the target programs. M. Clochard introduced one such framework, with an original extension that allows one to specify and prove fine-grained properties about infinite behaviors of target programs. The proof of correctness of this approach relies on a particular flavor of transfinite games. This proof is formalized and verified using the Why3 tool (http://toccata.lri.fr/gallery/hoare_logic_and_games.en.html).

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 [26]. 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 [128].

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 [24].

- **Verification of Parameterized Concurrent Programs on Weak Memory Models** Modern multiprocessors and microprocesseurs implement weak or relaxed memory models, in which the apparent order of memory operation does not follow the sequential consistency (SC) proposed by Leslie Lamport. Any concurrent program running on such architecture and designed with an SC model in mind may exhibit new behaviors during its execution, some of which may potentially be incorrect. For instance, a mutual exclusion algorithm, correct under an interleaving semantics, may no longer guarantee mutual exclusion when implemented on a weaker architecture. Reasoning about the semantics of such programs is a difficult task. Moreover, most concurrent algorithms are designed for an arbitrary number of processes. D. Declerck [12] proposed an approach to ensure the correctness of such concurrent algorithms, regardless of the number of processes involved. It relies on the Model Checking Modulo Theories (MCMT) framework, developed by Ghilardi and Ranise, which allows for the verification of safety properties of parameterized concurrent programs, that is to say, programs involving an arbitrary number of processes. This technology is extended with a theory for reasoning about weak memory models. The result is an extension of the Cubicle model checker called Cubicle-W, which allows the verification of safety properties of parameterized transition systems running under a weak memory model similar to TSO.
- **Counterexample Generation** S. Dailler and C. Marché worked on extensions and improvements of the counterexample generation feature of Why3, used in particular by the SPARK front-end for Ada [102] [101]. When the logic goal generated for a given verification condition is not shown unsatisfiable by an SMT solvers, some solver can propose a model. By carefully reverting the transformation chain (from an input program through the VC generator and the various translation steps to solvers), this model is turned into a potential counterexample that the user can exploit to analyze why its original code is not proved. The extension consists in a deep analysis of the complete model generated by the solver, so as to extract more information and produce better counterexamples. A journal paper giving the details of the whole process was published [14]
- **Alias Control for SPARK Program Verification** G.-A. Jaloyan and A. Paskevich, together with C. Dross, M. Maalej, and Y. Moy made a proposal for introduction of pointers to the SPARK language, based on permission-driven static alias analysis method inspired by Rust's borrow-checker and affine types [35]. By ensuring that at any point of execution any writable value can only be accessed through a single name, it is possible to apply the standard rules of Hoare logic (or weakest precondition calculus) to verify programs with pointers. The proposed framework was implemented in the GNAT Ada compiler and the SPARK toolset.

7.2. Automated Reasoning

- **A Why3 Framework for Reflection Proofs and its Application to GMP's Algorithms** Earlier works using Why3 showed that automatically verifying the algorithms of the arbitrary-precision integer library GMP exceeds the current capabilities of automatic solvers. To complete this verification, numerous cut indications had to be supplied by the user, slowing the project to a crawl. G. Melquiond and R. Rieu-Helf extended Why3 with a framework for proofs by reflection, with minimal impact on the trusted computing base. This framework makes it easy to write dedicated decision procedures that make full use of Why3's imperative features and are formally verified. This approach opens the way to efficiently tackling the further verification of GMP's algorithms [20], [27].
- **Expressive and extensible automated reasoning tactics for Coq** Proof assistants based on Type Theory, such as Coq, allow implementing effective automatic tactics based on computational reasoning (e.g. `lia` for linear integer arithmetic, or `ring` for ring theory). Unfortunately, these are usually limited to one particular domain. In contrast, SMTCoq is a modular and extensible tool, using external provers, which generalizes these computational approaches to combine multiple theories. It relies on a high-level interface, which offers a greater expressiveness, at the cost of more complex automation. Q. Garchery, in collaboration with C. Keller and V .Blot, designed two improvements to increase expressiveness of SMTCoq without impeding its modularity and

its efficiency: the first adds some support for universally quantified hypotheses, while the second generalizes the support for integer arithmetic to the different representations of natural numbers and integers in Coq. This work will be presented in the next JFLA [30]

Non-linear Arithmetic Reasoning for Control-Command Software State-of-the-art (semi-)decision procedures for non-linear real arithmetic address polynomial inequalities by mean of symbolic methods, such as quantifier elimination, or numerical approaches such as interval arithmetic. Although (some of) these methods offer nice completeness properties, their high complexity remains a limit, despite the impressive efficiency of modern implementations. This appears to be an obstacle to the use of SMT solvers when verifying, for instance, functional properties of control-command programs. Using off-the-shelf convex optimization solvers is known to constitute an appealing alternative. However, these solvers only deliver approximate solutions, which means they do not readily provide the soundness expected for applications such as software verification. S. Conchon, together with P. Roux and M. Iguernelala [21], investigated a-posteriori validation methods and their integration in the SMT framework. Although their early prototype, implemented in the Alt-Ergo SMT solver, often does not prove competitive with state of the art solvers, it already gives some interesting results, particularly on control-command programs.

Lightweight Interactive Proving for Automated Program Verification Deductive verification approach allows establishing the strongest possible formal guarantees on critical software. The downside is the cost in terms of human effort required to design adequate formal specifications and to successfully discharge the required proof obligations. To popularize deductive verification in an industrial software development environment, it is essential to provide means to progressively transition from simple and automated approaches to deductive verification. The SPARK environment, for development of critical software written in Ada, goes towards this goal by providing automated tools for formally proving that some code fulfills the requirements expressed in Ada contracts.

In a program verifier that makes use of automatic provers to discharge the proof obligations, a need for some additional user interaction with proof tasks shows up: either to help analyzing the reason of a proof failure or, ultimately, to discharge the verification conditions that are out-of-reach of state-of-the-art automatic provers. Adding interactive proof features in SPARK appears to be complicated by the fact that the proof toolchain makes use of the independent, intermediate verification tool Why3, which is generic enough to accept multiple front-ends for different input languages. S. Dailler, C. Marché and Y. Moy proposed an approach to extend Why3 with interactive proof features and also with a generic client-server infrastructure allowing integration of proof interaction into an external, front-end graphical user interface such as the one of SPARK. This was presented at the F-IDE symposium [18].

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 [90]. 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 [22].

Analysis of explicit Runge-Kutta methods Numerical integration schemes are mandatory to understand complex behaviors of dynamical systems described by ordinary differential equations. Implementation of these numerical methods involve floating-point computations and propagation of round-off errors. In the spirit of [58], S. Boldo, F. Faissole and A. Chapoutot developed a fine-grained analysis of round-off errors in explicit Runge-Kutta integration methods, taking into account exceptional behaviors, such as underflow and overflow [31].

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 [113]. 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 [15]. This makes CoqInterval one of the very few tools able to produce reliable results for improper integrals, be they formally verified or not.

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 [74]. An extended version that considers arbitrary rectangular matrices instead of square ones is published in the Journal of Automated Reasoning [13]. The development is available in Toccata's gallery http://toccata.lri.fr/gallery/verifythis_2016_matrix_multiplication.en.html.

Digital Filters Digital filters are small iterative algorithms, used as basic bricks in signal processing (filters) and control theory (controllers). D. Gallois-Wong, S. Boldo and T. Hilaire formally proved in Coq some error analysis theorems about digital filters, namely the Worst-Case Peak Gain theorem and the existence of a filter characterizing the difference between the exact filter and the implemented one. Moreover, as the digital signal processing literature provides many equivalent algorithms, called realizations, they formally defined and proved the equivalence of several realizations (Direct Forms and State-Space) [19]. Another Coq development dedicated the a realization called SIF (Specialized Implicit Form) has been done, in order to encompass all the other realizations up to the order of computation, which is very important in finite precision [25].

7.4. Floating-Point and Numerical Programs

Correct Average of Decimal Floating-Point Numbers Some modern processors include decimal floating-point units, with a conforming implementation of the IEEE-754 2008 standard. Unfortunately, many algorithms from the computer arithmetic literature are not correct anymore when computations are done in radix 10. This is in particular the case for the computation of the average of two floating-point numbers. S. Boldo, F. Faissole and V. Tourneur developed a new radix-10 algorithm that computes the correctly-rounded average, with a Coq formal proof of its correctness, that takes gradual underflow into account [17].

Optimal Inverse Projection of Floating-Point Addition In a setting where we have intervals for the values of floating-point variables x, a, and b, we are interested in improving these intervals when the floating-point equality $x \oplus a = b$ holds. This problem is common in constraint propagation, and called the inverse projection of the addition. It also appears in abstract interpretation for the analysis of programs containing IEEE 754 operations. D. Gallois-Wong, S. Boldo and P. Cuoq proposed floating-point theorems that provide optimal bounds for all the intervals. Fast loop-free algorithms compute these optimal bounds using only floating-point computations at the target precision [34].

Handbook of Floating-point Arithmetic Initially published in 2010, the *Handbook of Floating-Point Arithmetic* has been heavily updated. G. Melquiond contributed to the second edition [28].

Error analysis of finite precision digital filters and controllers The effort to provide accurate and reliable error analysis of fixed-point implementations of Signal Processing and Control algorithms was continued (see also the formalization effort above). A. Volkova, M. Istoan, F. de Dinechin and T. Hilaire (Citi Lyon, INSA Lyon) created an automatic code generator for FPGAs and dedicated roundoff analysis in order to minimize the bit-widths used for the intern computations while guaranteeing a bound on the output error [16]. The global workflow for the rigorous design of reliable Fixed-Point filters has been studied by A. Volkova, T. Hilaire and C. Lauter and submitted to a journal [36] : it concerns the rigorous determination of the Most Significant Bit of each variable, to guaranty that no overflow will ever occur, also taking into account the roundoff error propagation.

VERIDIS Project-Team

7. New Results

7.1. Automated and Interactive Theorem Proving

Participants: Jasmin Christian Blanchette, Martin Bromberger, Daniel El Ouraoui, Mathias Fleury, Pascal Fontaine, Stephan Merz, Hans-Jörg Schurr, Sorin Stratulat, Thomas Sturm, Andreas Teucke, Sophie Tourret, Marco Voigt, Uwe Waldmann, Christoph Weidenbach.

7.1.1. Extension of the Superposition Calculus with λ-free Higher-Order Terms and (Co)datatypes

Joint work with Alexander Bentkamp (VU Amsterdam), Simon Cruanes (Aesthetic Integration), Nicolas Peltier (IMAG Grenoble), and Simon Robillard (Chalmers Gothenburg).

Superposition is a highly successful calculus for reasoning about first-order logic with equality. As a stepping stone towards extending the calculus to full higher-order logic, Bentkamp et al. [19] designed a graceful generalization of the calculus to a fragment devoid of λ-abstractions, but with partial application and application of variables, two crucial higher-order features. This builds on the work on term orders, namely the recursive path order [57] and the Knuth-Bendix order [55]. We implemented the calculi in Simon Cruanes's Zipperposition prover and evaluated them on TPTP benchmarks. The performance is substantially better than with the traditional, encoding-based approach. The new superposition-like calculus serves as a stepping stone towards complete, efficient automatic theorem provers for full higher-order logic.

Another extension of superposition, by Blanchette et al. [21], concerns the native support for inductive and coinductive datatypes. The ability to reason about datatypes has many applications in program verification, formalization of the metatheory of programming languages, and even formalization of mathematics.

Both lines of work aim at bridging the gap between automatic and interactive theorem provers, by increasing the expressiveness and efficiency of best-of-breed automatic first-order provers based on the superposition calculus.

7.1.2. IsaFoL: Isabelle Formalization of Logic

Joint work with Alexander Bentkamp (VU Amsterdam), Andreas Halkjær From (DTU Copenhagen), Alexander Birch Jensen (DTU Copenhagen), Peter Lammich (TU München), John Bruntse Larsen (DTU Copenhagen), Julius Michaelis (TU München), Tobias Nipkow (TU München), Nicolas Peltier (IMAG Grenoble), Simon Robillard (Chalmers Gothenburg), Anders Schlichtkrull (DTU Copenhagen), Dmitriy Traytel (ETH Zürich), Jørgen Villadsen (DTU Copenhagen), and Petar Vukmirović (VU Amsterdam).

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.

The main result this year has been a formalization of a large part of Bachmair and Ganzinger's chapter on resolution theorem proving in the *Handbook of Automated Reasoning*, by Anders Schlichtkrull et al. The work was conducted by Schlichtkrull largely during a visit at the MPI in Saarbrücken and was published at IJCAR 2018 [34]. The following quote of one of the reviews nicely sums up the objective of the project:

[0] https://bitbucket.org/isafol/isafol/wiki/Home

> The authors convinced me that their development is a great tool for exploring/developing calculus extensions. It will enable us to "*extend/hack without fear*."

A follow-up paper [33], also by Schlichtkrull et al., has been accepted at CPP 2019. In this work, a chain of refinement leads to a verified executable prover.

The IsaFoL repository has welcome several further additions in 2018, and there is largely finished work, which we expect will lead to at least two publications in 2019:

- After the journal publication [13] following up on an IJCAR 2016 paper and a publication at CPP 2018 [23], Fleury has improved his verified SAT solver IsaSAT further by implementing four optimizations: restarts, forgetting, blocking literals, and machine integers. IsaSAT is now by far the most efficient verified SAT solver, and it is catching up with MiniSat, a reference (but unverified) SAT solver implementation.
- Sophie Tourret and Simon Robillard have formalized a new framework, designed primarily by Uwe Waldmann, that captures abstractly the lifting from completeness of a calculus for propositional logic to a first-order prover. This will yield a simpler proof of Bachmair and Ganzinger's completeness theorem and will be reusable for reasoning about other provers (e.g., superposition provers), whether with pen and paper or in Isabelle.

Jasmin Blanchette briefly describes this ongoing research in an invited paper [20], which he will present at CPP 2019.

7.1.3. Subtropical Reasoning for Real Inequalities

Joint work with Hoon Hong (North Carolina State University, Raleigh, NC).

We consider systems of strict multivariate polynomial inequalities over the reals. All polynomial coefficients are parameters ranging over the reals, where for each coefficient we prescribe its sign. We are interested in the existence of positive real solutions of our system for all choices of coefficients subject to our sign conditions. We give a decision procedure for the existence of such solutions. In the positive case our procedure yields a parametric positive solution as a rational function in the coefficients. Our framework allows heuristic subtropical approaches to be reformulated for non-parametric systems of polynomial inequalities. Such systems have been recently used in qualitative biological network analysis and, independently, in satisfiability modulo theory solving. We apply our results to characterize the incompleteness of those methods.

The approach allows SMT solving for non-linear real arithmetic to be heuristically reduced to linear real arithmetic, to which, e.g., methods from 7.1.4 are applicable. In the special case of single inequalities one can even reduce to linear programming. [25]. This has been successfully applied to heuristic search for Hopf bifurcation fixed points in chemical and biological network analysis.

7.1.4. Reasoning in Linear Arithmetic

We have continued our work on reasoning in linear integer (LIA), linear real (LRA) and linear mixed arithmetic (LIRA). Whereas the standard branch-and-bound techniques [63] for LIA typically work well for bounded systems of inequations, they often diverge on unbounded systems. We already proposed cube techniques for this case. They comprise efficiently computable sufficient tests for the existence of a solution [58]. However, these tests are only necessary for the existence of a solution in the case of a system that is unbounded in all directions. For the case of partially unbounded systems, our combination of the Mixed-Echelon-Hermite transformation and the Double-Bounded Reduction for systems of linear mixed arithmetic preserve satisfiability, can be computed in polynomial time, and turn any LIRA system into a bounded system [22]. Existing approaches for LIRA, e.g., branch-and-bound and cuts from proofs, only explore a finite search space after the application of our two transformations. The transformations orient themselves on the structure of an input system instead of computing *a priori* (over-)approximations out of the available constants. We also developed a polynomial method for converting certificates of (un)satisfiability from the transformed to the original system.

Meanwhile our techniques have been integrated into the SMT solver veriT, but also in other SMT solvers such as Z3 [72] or MathSAT [62]. They have been substantial for our success at SMTComp2018.

7.1.5. Combination of Satisfiability Procedures

Joint work with Christophe Ringeissen (Inria Nancy – Grand Est, Pesto) and Paula Chocron (IIIA-CSIC, Bellaterra, 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. 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 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 [60]. 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 [61] 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 2018, we have been improving the framework and unified both results. A paper is under review.

7.1.6. Quantifier Handling in SMT

Joint work with Andrew J. Reynolds (Univ. of Iowa, USA) and Cezary Kaliszyk (Univ. of Innsbruck).

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 in 2017 [53]. In a publication at TACAS 2018 [31], we revisit enumerative instantiation for SMT.

We are currently investigating machine learning techniques as a tool for filtering instantiations. Other ongoing work aims at lifting the above techniques to higher-order reasoning.

7.1.7. Real Quantifier Elimination, Decision, and Satisfiability and Their Applications

Effective quantifier elimination procedures for first-order theories provide a powerful tool for generically solving a wide range of problems based on logical specifications. In contrast to general first-order provers, quantifier elimination procedures are based on a fixed set of admissible logical symbols with an implicitly fixed semantics. This admits the use of sub-algorithms from symbolic computation. Specifically quantifier elimination for the reals has been successfully applied in geometry, verification, and the life sciences.

A survey paper with an invited talk at ISSAC 2018 provides a coherent view on the scientific developments of the virtual substitution method for real quantifier elimination during the past three decades [17]. Another recent survey paper had illustrated relevant applications of that method [71].

7.1.8. Non-Linear Arithmetic in SMT

Joint work with M. Ogawa and X. T. Vu (Japan Advanced Institute of Science and Technology), V. K. To (University of Engineering and Technology, VNU, Hanoi, Vietnam).

In the context of the SC^2 project (cf. sections 8.1 and 8.3), we study the theory, design techniques, and implement software to push forward the non-linear arithmetic (NLA) reasoning capabilities in SMT. Previously, 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. We also studied integration of these procedures into combinations of theories. These ideas were validated through an implementation within the veriT solver, together with code from the raSAT solver (from JAIST), and they were presented at the SC^2 workshop 2018 [24].

7.1.9. Proofs for SMT

We have previously 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 conference publication in 2017. In 2018, we polished the approach, fully implementing proof reconstruction of veriT proofs in Isabelle. A paper has been accepted in the Journal of Automated Reasoning.

7.1.10. A More Efficient Technique for Validating Cyclic Pre-Proofs

Cyclic pre-proofs can be represented as sets of finite tree derivations with back-links. In a setting of first-order logic with inductive definitions, the nodes of the tree derivations are labelled by sequents and the back-links connect particular terminal nodes, referred to as buds, to other nodes labelled by the same sequent. However, only some back-links can constitute sound pre-proofs. Previously, it was shown that special ordering and derivability conditions, defined along the minimal cycles of the digraph representing a particular normal form of the cyclic pre-proof, are sufficient for validating the back-links. In that approach, a single constraint could be checked several times when processing different minimal cycles, hence one may require additional recording mechanisms to avoid redundant computation in order to achieve polynomial time complexity.

In [39], we presented a new approach that does not need to process minimal cycles. It is based on a normal form in which the validation conditions are defined by taking into account only the root-bud paths from the non-singleton strongly connected components of its digraph.

7.1.11. Mechanical Synthesis of Algorithms by Logical and Combinatorial Techniques

Joint work with Isabela Dramnesc (West University, Timisoara, Romania) and Tudor Jebelean (RISC, Johannes Kepler University, Linz, Austria).

In [14], we developed logical and combinatorial methods for automating the generation of sorting algorithms for binary trees, starting from input-output specifications and producing conditional rewrite rules. The main approach consists in proving (constructively) the existence of an appropriate output from every input. The proof may fail if some necessary sub-algorithms are lacking. Then, their specifications are suggested and their synthesis is performed by the same principles.

The main goal is to avoid the possibly prohibitive cost of pure resolution proofs by using a natural-style proving in which domain-specific strategies and inference steps lead to a significant increase of efficiency. We introduce novel techniques and combine them with classical techniques for natural-deduction style proving, as well as methods based on the properties of domain-specific relations and functions. In particular, we use combinatorial techniques in order to generate possible witnesses, which in certain cases lead to the discovery of new induction principles. From the proof, the algorithm is extracted by transforming inductive proof steps into recursions, and case-based proof steps into conditionals.

The approach was demonstrated using the Theorema system for developing the theory, implementing the prover, and performing the proofs of the necessary properties and synthesis conjectures. It was also validated in the Coq system, allowing us to compare the facilities of the two systems in view of our application.

7.1.12. Formal Proofs of Tarjan's Algorithm

Joint work with Ran Chen (Chinese Academy of Sciences), Cyril Cohen and Laurent Théry (Inria Sophia Antipolis Méditerranée, Marelle), and Jean-Jacques Lévy (Inria Paris, Pi.r2).

We compare formal proofs of Tarjan's algorithm for computing strongly connected components in a graph in three different proof assistants: Coq, Isabelle/HOL, and Why3. Our proofs are based on a representation of the algorithm as a functional program (rather than its more conventional imperative representation), which was verified in Why3 by Chen and Lévy [59]. The proofs in all three assistants are thus closely comparable and in particular employ the same invariants. This lets us focus on different formalizations due to idiosyncracies of the proof assistants, such as w.r.t. handling mutually recursive function definitions whose termination is not obvious according to syntactic criteria, and compare the degree of automation in the three assistants. A report is available on arXiv [45].

7.2. Formal Methods for Developing and Analyzing Algorithms and Systems

Participants: Marie Duflot-Kremer, Yann Duplouy, Margaux Duroeulx, Souad Kherroubi, Igor Konnov, Dominique Méry, Stephan Merz, Axel Palaude, Nicolas Schnepf, Christoph Weidenbach.

7.2.1. Parameterized Verification of Threshold-Guarded Fault-Tolerant Distributed Algorithms

Joint work with Nathalie Bertrand (Inria Rennes, SUMO project team) and Jure Kukovec, Marijana Lazić, Ilina Stoilkovska, Josef Widder, Florian Zuleger (TU Wien).

Many fault-tolerant distributed algorithms use threshold guards: processes broadcast messages and count the number of messages that they receive from their peers. Based on the total number n of processes and an upper bound on the number t of faulty processes, a correct process tolerates faults by receiving "sufficiently many" messages. For instance, when a correct process has received $t+1$ messages from distinct processes, at least one of these messages must originate from a non-faulty process. The main challenge is to verify such algorithms for all combinations of parameters n and t that satisfy a resilience condition, e.g., $n > 3t$.

In earlier work, we introduced threshold automata for representing processes in such algorithms and showed that systems of threshold automata have bounded diameters that do not depend on the parameters such as n and t, provided that a single-step acceleration is allowed [66]. In the contribution [27] to CONCUR'18, we reported on various extensions of this result to less restrictive forms of automata: the guards can be non-linear, shared variables can be incremented and decremented, non-trivial loops are allowed, and more general forms of acceleration are used. In the contribution [26] to ISOLA'18, we presented a parallel extension of our tool Byzantine Model Checker (ByMC), which allows one to distribute the verification queries across the computation nodes in an MPI cluster.

Our previous results apply to asynchronous algorithms. It is well-known that distributed consensus cannot be solved in purely asynchronous systems [64]. However, when an algorithm is provided with a random coin, consensus becomes solvable [56]. In [44], we introduced an approach to parameterized verification of randomized threshold-guarded distributed algorithms, which proceed in an unbounded number of rounds and toss a coin to break symmetries. This approach integrates two levels of reasoning: (1) proving safety and liveness of a single round system with ByMC by replacing randomization with non-determinism, (2) showing almost-sure termination of an algorithm by using the verification results for the non-deterministic system. To show soundness, we proved several theorems that reduce reasoning about multiple rounds to reasoning about a single round. We verified five prominent algorithms, including Ben-Or's randomized consensus [56] and randomized one-step consensus (RS-BOSCO [70]). The verification of the latter algorithm required us to run experiments in Grid5000. A paper describing these results is under review at TACAS 2019.

Another way of making consensus solvable is to impose synchrony on the executions of a distributed system. In [48] we introduced synchronous threshold automata, which execute in lock-step and count the number of processes in given local states. In general, we showed that even reachability of a parameterized set of global states in such a distributed system is undecidable. However, we proved that systems of automata with monotonic guards have bounded diameters, which allows us to use SMT-based bounded model checking as a complete parameterized verification technique. We introduced a procedure for computing the diameter of a counter system of synchronous threshold automata, applied it to the counter systems of 8 distributed algorithms from the literature, and found that their diameters are tiny (from 1 to 4). This makes our approach practically feasible, despite undecidability in general. A paper about this work is under review at TACAS 2019.

7.2.2. Symbolic Model Checking of TLA+ Specifications

Joint work with Jure Kukovec, Thanh Hai Tran, Josef Widder (TU Wien).

TLA$^+$ is a general language introduced by Leslie Lamport for specifying temporal behavior of computer systems [67]. The tool set for TLA$^+$ includes an explicit-state model checker TLC. As explicit state model checkers do not scale to large verification problems, we started the project APALACHE [0] on developing a symbolic model checker for TLA$^+$ in 2016.

In the contribution [28] to ABZ'18, we addressed the first principal challenge towards developing the symbolic model checker. We introduced a technique for identifying assignments in TLA$^+$ specifications and decomposing a monolithic TLA$^+$ specification into a set of symbolic transitions. At the TLA$^+$ community meeting 2018, we presented a prototype solution [46] to a second challenge. We have developed an SMT encoding of TLA$^+$ expressions for model checking purposes. We presented the first version of a symbolic model checker for TLA$^+$ specifications that works under the same assumptions as TLC: the input parameters are fixed and finite structures, and the reachable states are finite structures. The experimental results are encouraging, and we are thus preparing a conference submission. Finally, in a contribution to the DSN Workshop on Byzantine Consensus and Resilient Blockchains [47], we considered challenges for automatic verification techniques for Blockchain protocols.

7.2.3. Making Explicit Domain Knowledge in Formal System Development

Joint work with partners of the IMPEX project.

The IMPEX project (cf. section 8.1) advocates 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. We strive at offering rigorous mechanisms for handling domain knowledge in design models. The main results of the project are summarized in [18] and show the importance of three operations over models, namely annotation, dependency, and refactoring [38].

7.2.4. 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 about the conformance of each refinement with the model preceding it. Our main result during 2018 is the development of patterns for different kinds of paradigms including the iterative pattern, the recursive pattern, and the distributed pattern [30].

7.2.5. Synthesis of Security Chains for Software Defined Networks

Joint work with Rémi Badonnel and Abdelkader Lahmadi of the Resist team of Inria Nancy – Grand Est.

The PhD work of Nicolas Schnepf focuses on applying formal methods techniques in the area of network communications, and in particular for the construction, analysis, and optimization of security functions in the setting of software-defined networks (SDN). In previous work, we defined an extension of the Pyretic language [65] for representing 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.

[0] WWTF project APALACHE (ICT15-103): https://forsyte.at/research/apalache/

This year, our work focused on synthesizing security chains for Android applications based on their observed communications. The first step consists in inferring probabilistic finite-state automata models that represent network flows generated by Android applications. Comparing our models with automata produced by the state-of-the-art tools Invarimint and Synoptic, we obtain representations that are significantly smaller than those generated by Synoptic and as succinct as those inferred by Invarimint, but that include information about transition probability, unlike Invarimint. This work was presented at NOMS 2018 [35], [37]. In a second step, we encode security policies defined by network administrators in a rule-based program that is then used to generate a high-level representation of a security chain for the application, which is then translated to Pyretic. For example, an application that contacts different ports at the same IP address in rapid succession could be qualified as performing a port scanning attack, and these connections could then be blocked. This work was presented at AVoCS 2018 [36]. The third step consists in factorizing the chains generated for different applications in order to reduce the size of the overall chain that must be deployed in a network. A paper describing appropriate algorithms for that purpose will be presented at IM 2019.

7.2.6. 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 traditionally 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 2018, our work focused on the encoding of dynamic fault trees whose structure function needs to take into account the order in which components fail.

CIDRE Project-Team

6. New Results

6.1. Axis 1 : Attack comprehension

6.1.1. Attacks stay possible even when programs seem not vulnerable

The protection of any software starts at the hardware level. In ,K. Bukasa, L. Claudepierre, J.-L. Lanet, in collaboration with R. Lashermes from SED Inria Rennes – Bretagne Atlantique, explore how Electromagnetic Fault Injection (EMFI) can disturb the behavior of a chip and undermine the security of the information handled by the target. They demonstrate the possibilities to create software vulnerabilities with hardware fault injection (with EM pulses), not against crypto-systems but targeting regular software running on IoT devices. Experimentations are conducted on an ARMv7-M (Cortex-M3) microcontroller, present at the heart of a wide-range of embedded systems, to prove that a fault attack is able to create a vulnerability in a code where there is none in the usual software security meaning. Protecting against vulnerabilities must thus encompass protecting against both software and hardware attacks.

6.2. Axis 2 : Attack detection

6.2.1. Intrusion detection in sequential control systems.

Sophisticated process-aware attacks targeting industrial control systems require adequate detection measures taking into account the physical process. In [20], 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 the mining efficiency and attack detection capabilities of our approach.

6.2.2. Hardware-based Information Flow Tracking

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.

Existing hardware DIFT approaches have not been widely used neither by research community nor by hardware vendors. It is due to two major reasons: current hardware DIFT solutions lack support for multi-threaded applications and implementations for hardcore processors. In [10] we addresse both issues by introducing an approach with some unique features: DIFT for multi-threaded software, virtual memory protection (rather than physical memory as in related works) and Linux kernel support using an information flow monitor called RFBlare. These goals are accomplished by taking advantage of a notable feature of ARM CoreSight components (context ID) combined with a custom DIFT coprocessor and RFBlare. The communication time overhead, major source of slowdown in total DIFT time overhead, is divided by a factor 3.8 compared to existing solutions with similar software constraints as in this work. The area overhead of this work is lower than 1% and power overhead is 16.2% on a middle-class Xilinx Zynq SoC.

Most of hardware-assisted solutions for software security, program monitoring, and event-checking approaches require instrumentation of the target software, an operation which can be performed using an SBI (Static Binary Instrumentation) or a DBI (Dynamic Binary Instrumentation) framework. Hardware-assisted instrumentation can use one of these two solutions to instrument data to a memory-mapped register. Both these approaches require an in-depth knowledge of frameworks and an important amount of software modifications in order to instrument a whole application. In [11] we propose a novel way to instrument an application, at the source code level, taking advantage of underlying hardware debug components such as CS (CoreSight) components available on Xilinx Zynq SoCs. As an example, the instrumentation approach proposed in this work is used to detect a double free security attack. Furthermore, it is evaluated in terms of runtime and area overhead.

6.2.3. Alert correlation in intrusion detection.

In distributed systems and in particular in industrial SCADA environments, alert correlation systems are necessary to identify complex multi-step attacks within the huge amount of alerts and events. In [22] we describe an automata-based correlation engine developed in the context of a European project where the main stakeholder was an energy distribution company. The behavior of the engine is extended to fit new requirements. In the proposed solution, a fully automated process generates thousands of correlation rules. Despite this major scalability challenge, the designed correlation engine exhibits good performance. Expected rates of incoming low level alerts approaching several hundreds of elements per second are tolerated. Moreover, the data structures chosen allow to quickly handle dynamic changes of the set of correlation rules. As some attack steps are not observed, the correlation engine can be tuned to raise an alert when all the attack steps except k of them have been detected. To be able to react to an ongoing attack by taking countermeasures, alerts must also be raised as soon as a significant prefix of an attack scenario is recognized. Fulfilling these additional requirements leads to an increase in the memory consumption. Therefore purge mechanisms are also proposed and analyzed. An evaluation of the tool is conducted in the context of a SCADA environment.

6.2.4. Most recent and frequent items in distributed streams for DDoS detection.

The need to analyze in real time large-scale and distributed data streams has recently became tremendously important to detect attacks (DDoS), anomalies or performance issues. In particular the identification of recent heavy-hitters (or hot items) is essential but highly challenging. Actually, this problem has been heavily studied during the last decades with both exact and probabilistic solutions. While simple to state and fundamental for advanced analysis, answering this issue over a sliding time window and among distributed nodes is still an active research field. The distributed detection of frequent items over a sliding time window presents two extra challenging aspects with respect to the centralized detection of frequent items since the inception of the stream: (i) Treat time decaying items as they enter and exit the sliding window; (ii) Produce mergeable local stream summaries in order to obtain a system-wide summary. In [12], we propose a sliding window-based solution of the top k most frequent items based on a deterministic counting of the most over-represented items in the data streams, which are themselves probabilistically identified using a dynamically defined threshold. Performance of our new algorithm are astonishingly good, despite any items order manipulation or distributed execution.

6.2.5. Propagation of information.

Together with Yves Mocquard and Bruno Sericola, we have worked on the well studied dissemination of information in large scale distributed networks through pairwise interactions. The information to be propagated can simply be a bit of information to any code, including viruses. This problem, originally called rumor mongering, and then rumor spreading has mainly been investigated in the synchronous model. This model relies on the assumption that all the nodes of the network act in synchrony, that is, at each round of the protocol, each node is allowed to contact a random neighbor. In this paper, we drop this assumption under the argument that it is not realistic in large scale systems. We thus consider the asynchronous variant, where at random times, nodes successively interact by pairs exchanging their information on the rumor. In a previous paper, we performed a study of the total number of interactions needed for all the nodes of the network to discover the rumor. While most of the existing results involve huge constants that do not allow us to compare different protocols, we provided a thorough analysis of the distribution of this total number of interactions together with its asymptotic behavior [4]. In addition to this study, we have proposed an algorithm that allows, through simple pairwise interactions, each node of the large scale and dynamic system to build a global clock which allows any node to maintain with high probability a common temporal referential [25]. By combining this global clock together with the rumor spreading algorithm, we have proposed a mechanism that allows each node to locally detect that the system has converged to a sought configuration with high probability. We have also shown the applicability of our convergence detection mechanism to many other pairwise interaction-based protocols. For instance, our construction can be applied to a leader election protocol provided that its convergence time is known with high probability [26].

6.3. Axis 3 : Attack resistance

6.3.1. Connectivity in an inter-MANET network.

New generation radio equipment, used by soldiers and vehicles on the battlefield, form ad hoc networks and specifically, Mobile Ad hoc NETworks (MANET). The battlefields where these equipments are deployed include a majority of coalition communication. Each group on the battleground may communicate with other members of the coalition and establish inter-MANET links. These inter-MANET links are governed by routing policies that can be summarized as Allowed or Denied link. However, if more than two groups form a coalition, blocked multi-hop communications and non-desired transmissions due to these restrictive policies would appear. In [19], we present these blocking cases and theoretically evaluate their apparition frequency. Then, we present two alternatives to extend the binary policies and decrease the number of blocking cases. Finally, we describe an experimental scenario containing a blocking case and evaluate our propositions and their performance.

6.3.2. Permissionless ledgers for decentralized cryptocurrency systems (blockchain).

The goal of decentralized cryptocurrency systems is to offer a medium of exchange secured by cryptography, without the need of a centralized banking authority. An increasing number of distributed cryptocurrency systems are emerging, and among them Bitcoin, which is often designated as the pioneer of this kind of systems. Bitcoin circumvents the absence of a global trusted third-party by relying on a blockchain, an append-only data-structure, publicly readable and writable, in which all the valid transactions ever issued in the system are progressively appended through the creation of cryptographically linked blocks. In [15], we propose a new way to organise both transactions and blocks in a distributed ledger to address the performance issues of permissionless ledgers. In contrast to most of the existing solutions in which the ledger is a chain of blocks extracted from a tree or a graph of chains, we present a distributed ledger whose structure is a balanced directed acyclic graph of blocks. We call this specific graph a SYC-DAG. We show that a SYC-DAG allows us to keep all the remarkable properties of the Bitcoin blockchain in terms of security, immutability, and transparency, while enjoying higher throughput and self-adaptivity to transactions demand.

6.3.3. Modular verification of Programs with Effects and Effect Handlers in Coq

Modern computing systems have grown in complexity, and the attack surface has increased accordingly. Even though system components are generally carefully designed and even verified by different groups of people, the composition of these components is often regarded with less attention. This paves the way for architectural attacks, a class of security vulnerabilities where the attacker is able to threaten the security of the system even if each of its components continues to act as expected. In [24], we introduce FreeSpec, a formalism built upon the key idea that components can be modelled as programs with algebraic effects to be realized by other components. FreeSpec allows for the modular modelling of a complex system, by defining idealized components connected together, and the modular verification of the properties of their composition. In addition, we have implemented a framework for the Coq proof assistant based on FreeSpec.

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. Secure Information Flow and Game Theory

In the inference attacks studied in Quantitative Information Flow (QIF), the attacker typically tries to interfere with the system in the attempt to increase its leakage of secret information. The defender, on the other hand, typically tries to decrease leakage by introducing some controlled noise. This noise introduction can be modeled as a type of protocol composition, i.e., a probabilistic choice among different protocols, and its effect on the amount of leakage depends heavily on whether or not this choice is visible to the attacker. In [21], [11], we considered operators for modeling visible and hidden choice in protocol composition, and we studied their algebraic properties. We then formalized the interplay between defender and attacker in a game-theoretic framework adapted to the specific issues of QIF, where the payoff is information leakage. We considered various kinds of leakage games, depending on whether players act simultaneously or sequentially, and on whether or not the choices of the defender are visible to the attacker. In the case of sequential games, the choice of the second player is generally a function of the choice of the first player, and his/her probabilistic choice can be either over the possible functions (mixed strategy) or it can be on the result of the function (behavioral strategy). We showed that when the attacker moves first in a sequential game with a hidden choice, then behavioral strategies are more advantageous for the defender than mixed strategies. This contrasts with the standard game theory, where the two types of strategies are equivalent. Finally, we established a hierarchy of these games in terms of their information leakage and provide methods for finding optimal strategies (at the points of equilibrium) for both attacker and defender in the various cases.

6.1.2. The additive capacity problem for Quantitative Information Flow

Preventing information leakage is a fundamental goal in achieving confidentiality. In many practical scenarios, however, eliminating such leaks is impossible. It becomes then desirable to quantify the severity of such leaks and establish bounds on the threat they impose. Aiming at developing measures that are robust wrt a variety of operational conditions, a theory of channel capacity for the g-leakage model was developed in [25], providing solutions for several scenarios in both the multiplicative and the additive setting. In [16] we continued this line of work by providing substantial improvements over the results of [25] for additive leakage. The main idea of employing the Kantorovich distance remains, but it is now applied to quasimetrics, and in particular the novel "convex-separation" quasimetric. The benefits were threefold: first, it allowed to maximize leakage over a larger class of gain functions, most notably including the one of Shannon. Second, a solution was obtained to the problem of maximizing leakage over both priors and gain functions, left open in [25]. Third, it allowed to establish an additive variant of the " Miracle " theorem from [26].

6.1.3. Local Differential Privacy and Statistical Utility

Local differential privacy (LDP) is a variant of differential privacy (DP) where the noise is added directly on the individual records, before being collected. The main advantage with respect to DP is that we do not need a trusted third party to collect and sanitise the sensitive data of the user. The main disadvantage is that the trade-off between privacy and utility is usually worse than in DP, and typically to retrieve reasonably good statistics from the locally sanitised data it is necessary to have access to a huge collection of them. In [22], we focused on the problem of estimating the counting queries on numerical data, and we proposed a variant of LDP based on the addition of geometric noise. Such noise function is known to have appealing properties in the case of counting queries. In particular, it is universally optimal for DP, i.e., it provides the best utility for a given level of DP, regardless of the side knowledge of the attacker. We explored the properties of geometric noise for counting queries in the LDP setting, and we conjectured an optimality property, similar to the one that holds in the DP setting. In [15] we proposed a variant of LDP suitable for metric spaces, such as location data or energy consumption data, and we showed that it provides a better utility, for the same level of privacy, then the other known LPD mechanisms.

6.1.4. Information-Theoretic Methods for Feature Selection in Machine Learning

The identification of the "best" features for classification is a problem of increasing importance in machine learning. The size of available datasets is becoming larger and larger, both in terms of samples and in terms of features of the samples, and keeping the dimensionality of the data under control is necessary for avoiding an explosion of the training complexity and for the accuracy of the classification. The known methods for reducing the dimensionality can be divided in two categories: those which transform the feature space by reshaping the original features into new ones (feature extraction), and those which select a subset of the features (feature selection). Several proposals for feature selection have successfully applied concepts and techniques from information theory. In [19] we proposed a new information-theoretic algorithm for ordering the features according to their relevance for classification. The novelty of our proposal consisted in adopting Rényi min-entropy instead of the commonly used Shannon entropy. In particular, we adopted a notion of conditional min-entropy that has been recently proposed in the field of security and privacy, and that avoids the anomalies of previously-attempted information-theoretic definitions. This notion is strictly related to the Bayes error, which is a promising property for achieving accuracy in the classification. We evaluated our method on various classifiers and datasets, and we showed that it compares favorably to the corresponding one based on Shannon entropy.

6.1.5. A Logical Characterization of Differential Privacy via Behavioral Metrics

Differential privacy (DP) is a formal definition of privacy ensuring that sensitive information relative to individuals cannot be inferred by querying a database. In [18], we exploited a modeling of this framework via labeled Markov Chains (LMCs) to provide a logical characterization of differential privacy: we considered a probabilistic variant of the Hennessy-Milner logic and we defined a syntactical distance on formulae in it measuring their syntactic disparities. Then, we defined a trace distance on LMCs in terms of the syntactic distance between the sets of formulae satisfied by them. We proved that such distance corresponds to the level of privacy of the LMCs. Moreover, we used the distance on formulae to define a real-valued semantics for them, from which we obtained a logical characterization of weak anonymity: the level of anonymity is measured in terms of the smallest formula distinguishing the considered LMCs. Then, we focused on bisimulation semantics on nondeterministic probabilistic processes and we provide a logical characterization of generalized bisimulation metrics, namely those defined via the generalized Kantorovich lifting. Our characterization is based on the notion of mimicking formula of a process and the syntactic distance on formulae, where the former captures the observable behavior of the corresponding process and allows us to characterize bisimilarity. We showed that the generalized bisimulation distance on processes is equal to the syntactic distance on their mimicking formulae. Moreover, we used the distance on mimicking formulae to obtain bounds on differential privacy.

6.1.6. Probability and Nondeterminism in Process Calculi from a Logical Perspective

Behavioral equivalences and modal logics have been successfully employed for the specification and verification of communicating concurrent systems, henceforth processes. The former ones, in particular the family of bisimulations, provide a simple and elegant tool for the comparison of the observable behavior of processes. The latter ones allow for an immediate expression of the desired properties of processes. Since the work on the Hennessy-Milner logic (HML), these two approaches are connected by means of logical characterizations of behavioral equivalences: two processes are behaviorally equivalent if and only if they satisfy the same formulae in the logic. Hence, the characterization of an equivalence subsumes both the fact that the logic is as expressive as the equivalence and the fact that the equivalence preserves the logical properties of processes.

However, the connection between behavioral equivalences and modal logics goes even further: modal decomposition of formulae exploits the characterization of an equivalence to derive its compositional properties. Roughly speaking, the definition of the semantic behavior of processes by means of the Structural Operational Semantics (SOS) framework allowed for decomposing the satisfaction problem of a formula for a process into the verification of the satisfaction problem of certain formulae for its subprocesses. In [12] we extended the SOS-driven decomposition approach to processes in which the nondeterministic behavior coexists with probability. To deal with the probabilistic behavior of processes, and thus with the decomposition of formulae characterizing it, we introduced a SOS-like machinery allowing for the specification of the behavior of open distribution terms. By our decomposition, we obtained (pre)congruence formats for probabilistic bisimilarity, ready similarity and similarity.

The combination of nondeterminism and probability in concurrent systems leads to different interpretations of process behavior. If we restrict our attention to linear properties only, we can identify three main approaches to trace and testing semantics: the trace distributions, the trace-by-trace and the extremal probabilities approaches. In [17] we proposed novel notions of behavioral metrics that are based on the three classic approaches above, and that can be used to measure the disparities in the linear behavior of processes wrt. trace and testing semantics. We studied the properties of these metrics, like non-expansiveness, and we compare their expressive powers.

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. Real-time Rewriting Logic Semantics for Spatial Concurrent Constraint Programming

In [20] we used rewriting logic for specifying and analyzing a calculus for concurrent constraint programming (ccp) processes combining spatial and real-time behavior. These processes can run processes in different computational spaces (e.g., containers) while subject to real-time requirements (e.g., upper bounds in the execution time of a given operation), which can be specified with both discrete and dense linear time. The real-time rewriting logic semantics is fully executable in Maude with the help of rewriting modulo SMT: partial information (i.e., constraints) in the specification is represented by quantifier-free formulas on the shared variables of the system that are under the control of SMT decision procedures. The approach is used to symbolically analyze existential real-time reachability properties of process calculi in the presence of spatial hierarchies for sharing information and knowledge.

6.2.2. Characterizing Right Inverses for Spatial Constraint Systems with Applications to Modal Logic

In [14] spatial constraint systems are 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 is identified and the abstract notion of normality is shown to correspond to the preservation of finite suprema. Furthermore, a taxonomy of normal right inverses is 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 [13] 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.

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 [9], we showed how the movement above ground of the intermediation activity transforms territories. A global analysis of the geopolitics of technology was presented in [3].

7.2. Anthropocene studies

We have 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. More comprehensive approaches that rely on digital systems to control society and nudge citizens to adapt their behavior have been developed in Asia. We analyse in particular the social scoring system in China, and Society 5.0 in Japan [6]. An investigation of the world of images and photography in the time of algorithms was conducted in [2].

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 [1], we suggest that a new framework is necessary to apprehend new phenomena, such as 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 [5] a packet capture archive system is developed and described. In [4] 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 [7].

PESTO Project-Team

7. New Results

7.1. Security protocols

7.1.1. 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). A wide range of security properties, such as anonymity properties in electronic voting and auctions, unlinkability in RFID protocols and mobile phone protocols, are however naturally expressed in terms of indistinguishability, which is not a trace property. Indistinguishability is naturally formalized as an observational or trace equivalence in cryptographic process calculi, such as the applied pi calculus. While several decision procedures have already been proposed for verifying equivalence properties the resulting tools are often rather limited, and lack efficiency.

Our results are centered around the development of several, complementary verification tools for verifying equivalence properties. These tools are complementary in terms of expressivity, precision and efficiency.

- The *Akiss* tool provides good expressivity as it supports a large number of cryptographic primitives (including the XOR primitive, extremely popular in low energy devices such as RFID tags) and protocols with else branches. It allows verification for a bounded number of protocol sessions. The tool is precise for a class of determinate processes, and can approximate equivalence for other protocols. The tool however suffers from efficiency problems when the number of sessions increases. The computation can be partially distributed on different cores. To overcome these efficiency problems of the *Akiss* tool, Gazeau and Kremer completely revisit the theory underlying *Akiss*. Rather than enumerating the possible traces, the new version directly reasons about partial ordered traces. A new implementation is also in progress and the first results seem extremely promising.
- The SAT-Equiv tool is based on a novel algorithm, based on graph planning and SAT-solving. The tool has a limited expressivity in that it allows only the most standard cryptographic primitives, requires protocols to be determinate and does not support protocols with else branches. The tool is however extremely efficient, allowing verification for a very large (but bounded) number of sessions (where most other tools have to stop after one or two sessions). Cortier and Dallon, in collaboration with Delaune (IRISA), have presented at ESORICS'18 [20] an extension of SAT-EQUIV to support protocols with phases and a large class of cryptographic primitives that encompasses standard primitives. This required to first show a small attack property: whenever two protocols are not in equivalence, there exists a well-typed witness of non equivalence. This result was initially proved for symmetric encryption only and now holds for a large class of primitives [37].
- The DEEPSEC tool, presented by Cheval, Kremer and Rakotonirina at S&P'18 [18], is a new tool that allows for user-defined cryptographic primitives that can be modelled as a subterm convergent rewrite system (slightly more restricted than AKISS), but supports the whole applied pi calculus, except for bounding the number of sessions. It is precise, in that it decides equivalence (without any approximations) and has good efficiency (slightly less than SAT-Equiv) for the class of determinate processes (where partial order reductions apply). Their work also settled the question of the exact complexity of deciding different equivalences - static equivalence, trace equivalence and bisimulation. In particular they were able to show that both deciding trace equivalence and bisimulation in the case of cryptographic primitives modelled by subterm convergent rewrite systems are co-NEXP complete problems – this is a strong, new insight, solving a longstanding open question about the complexity of this problem. The DEEPSEC tool also implements state of the art partial order reductions and the verification can be distributed on different cores on a single machine and also on clusters of machines, as detailed in a CAV'18 tool paper [19].

- Unlike the above tools, the TYPE-EQ tool supports verification of both a bounded and unbounded number of protocol sessions (and a mix of them). It is based on a novel approach for equivalence properties. Instead of *deciding* equivalence like for the previous approaches, the tool uses a type system which is sound w.r.t. equivalence. Regarding precision, the tool is not complete, i.e. it may provide false attacks. It induces a significant speedup compared to previous tools for a bounded number of sessions and compares similarly to ProVerif [47] for an unbounded number of sessions. In collaboration with Maffei and Grimm, Lallemand and Cortier [23] extend this approach to all standard primitives and improve its precision, allowing to branch on secrets.

From a more foundational point of view, Ringeissen, in collaboration with Erbatur (LMU, Germany) and Marshall (Univ Mary Washington, USA), study decision procedures for two knowledge problems critical to the verification of security protocols, namely the intruder deduction and the static equivalence problems. These problems can be related to particular forms of context matching and context unification. Both problems are defined with respect to an equational theory and are known to be decidable when the equational theory is given by a subterm convergent term rewrite system. In a paper presented at UNIF'18 [33] they investigate the case of a subterm convergent equational term rewrite system defined modulo an equational theory, like Commutativity or Associativity-Commutativity. They show that for certain classes of such equational theories, namely the shallow classes, the two knowledge problems remain decidable.

7.1.2. Verification of protocols with global states

Participants: Vincent Cheval, Véronique Cortier, Jannik Dreier, Mathieu Turuani.

One known challenge when analysing security protocols for an unbounded number of sessions is the case of protocols with global states such as counters, tables, or more generally, memory cells. The popular tool ProVerif [47] fails to analyse such protocols, due to its internal abstraction. Cheval, Cortier, and Turuani have devised a generic transformation of the security properties queried to ProVerif. In a paper presented at CSF'18 [17], they proved the soundness of the transformation and implement it into a front-end GSVerif. Their experiments show that GSVerif (combined with ProVerif) outperforms the few existing tools, both in terms of efficiency and protocol coverage. GSVerif was successfully applied to a dozen of protocols of the literature, yielding the first fully automatic proof of a security API and a payment protocol of the literature.

The *TAMARIN* prover is a state-of-the-art verification tool for cryptographic protocols in the symbolic model. Dreier, in collaboration with Hirschi, Sasse (ETH Zurich), and Radomirovic (Dundee), improved the underlying theory and the tool to deal with an equational theory modeling XOR operations. Exclusive-or (XOR) operations are common in cryptographic protocols, in particular in RFID protocols and electronic payment protocols. Although there are numerous applications, due to the inherent complexity of faithful models of XOR, there is only limited tool support for the verification of cryptographic protocols using XOR. This makes *TAMARIN* the first tool to support simultaneously this large set of equational theories, protocols with global mutable state, an unbounded number of sessions, and complex security properties including observational equivalence. We demonstrated the effectiveness of our approach by analyzing several protocols that rely on XOR, in particular multiple RFID-protocols, where we can identify attacks as well as provide proofs. These results were presented at CSF'18 [29].

7.1.3. Analysis of deployed protocols

Participants: Jannik Dreier, Charlie Jacomme, Steve Kremer.

7.1.3.1. Multi-factor authentication.

Passwords are still the most widespread means for authenticating users, even though they have been shown to create huge security problems. This motivated the use of additional authentication mechanisms used in so-called multi-factor authentication protocols. In a paper, published at CSF'18 [30] Jacomme and Kremer define a detailed threat model for this kind of protocols: while in classical protocol analysis attackers control the communication network, the idea is to take into account that many communications are performed over TLS channels, that computers may be infected by different kinds of malwares, that attackers could perform phishing, and that humans may omit some actions. This model has been formalized in the applied pi calculus and perform an extensive analysis and comparison of several widely used protocols — variants of Google 2 step and FIDO U2F. The analysis is completely automated, generating systematically all combinations of threat scenarios for each of the protocols and using the ProVerif tool [47] for automated protocol analysis. Even though threat scenarios are eliminated as soon as results are implied by weaker scenarios, the analysis required over 6 000 calls to ProVerif, yet finishes in only a few minutes. Their analysis highlights weaknesses and strengths of the different protocols, and allows them to suggest several small modifications of the existing protocols which are easy to implement, yet improve their security in several threat scenarios.

7.1.3.2. 5G Authentication.

Mobile communication networks connect much of the world's population. The security of users' calls, SMSs, and mobile data depends on the guarantees provided by the Authenticated Key Exchange protocols used. For the next-generation network (5G), the 3GPP group has standardized the 5G AKA protocol for this purpose. We provided the first comprehensive formal model of a protocol from the AKA family: 5G AKA. We also extracted precise requirements from the 3GPP standards defining 5G and we identified missing security goals. Using the security protocol verification tool Tamarin and its recent extension to support XOR, we conducted a full, systematic, security evaluation of the model with respect to the 5G security goals. Our automated analysis identifies the minimal security assumptions required for each security goal and we found that some critical security goals are not met, except under additional assumptions missing from the standard. Finally, we made explicit recommendations with provably secure fixes for the attacks and weaknesses we found. These results were presented at CCS'18 [13].

7.1.3.3. Authentication Methods with PIN Codes.

Touch screens have become ubiquitous in the past few years, like for instance in smartphones and tablets. These devices are often the entry door to numerous information systems, hence having a secure and practical authentication mechanism is crucial. In this work, we examined the complexity of different authentication methods specifically designed for such devices. We studied the common technology to authenticate a user using a Personal Identifier Number code (PIN code). Entering the code is a critical moment where there are several possibilities for an attacker to discover the secret. We considered three attack models: a Bruteforce Attack (BA) model, a Smudge Attack (SA) model, and an Observation Attack (OA) model where the attacker sees the user logging in on his device. The aim of the intruder is to learn the secret code. Our goal is to propose alternative methods to enter a PIN code. We compared such different methods in terms of security. Some methods require more intentional resources than other, this is why we performed a psychological study on the different methods to evaluate the users' perception of the different methods and their usage. This work was presented at RCIS'18 [16].

7.1.4. Protocol design

Participant: Jannik Dreier.

7.1.4.1. A Cryptographer's Conspiracy Santa.

In Conspiracy Santa, a variant of Secret Santa, a group of people offer each other Christmas gifts, where each member of the group receives a gift from the other members of the group. To that end, the members of the group form conspiracies, to decide on appropriate gifts, and usually divide the cost of the gift among all participants of the conspiracy. This requires to settle the shared expenses per conspiracy, so Conspiracy Santa can actually be seen as an aggregation of several shared expenses problems. In this work, we showed that the problem of finding a minimal number of transactions when settling shared expenses is NP-complete. Still, there exists good greedy approximations. Second, we presented a greedy distributed secure solution to Conspiracy Santa. This solution allows a group of people to share the expenses for the gifts in such a way that no participant will learn the price of his/her gift, but at the same time notably reduces the number of transactions with respect to a naive aggregation. Furthermore, our solution does not require a trusted third party, and can either be implemented physically (the participants are in the same room and exchange money) or, virtually, using a cryptocurrency. This work was presented at FUN'18 [14].

7.1.4.2. A Physical Zero-Knowledge Proof for Makaro.

Makaro is a logic game similar to Sudoku. In Makaro, a grid has to be filled with numbers such that: given areas contain all the numbers up to the number of cells in the area, no adjacent numbers are equal, and some cells provide restrictions on the largest adjacent number. In this work we proposed a proven secure physical algorithm, only relying on cards, to realize a zero-knowledge proof of knowledge for Makaro. It allows a player to show that he/she knows a solution without revealing it. This work was presented at SSS'18 [15].

7.2. E-voting

7.2.1. Definitions for e-voting

Participants: Sergiu Bursuc, Véronique Cortier, Steve Kremer, Joseph Lallemand.

Electronic voting typically aims at two main security goals: vote privacy and verifiability. Verifiability typically includes individual verifiability (a voter can check that his/her ballot is counted); universal verifiability (anyone can check that the result corresponds to the published ballots); and eligibility verifiability (only legitimate voters may vote). Cortier and Lallemand have shown that privacy actually implies individual verifiability. In other words, systems without individual verifiability cannot achieve privacy (under the same trust assumptions). To demonstrate the generality of the result, they show this implication in two different settings, namely cryptographic and symbolic models, for standard notions of privacy and individual verifiability. This also highlights limitations in existing privacy definitions in cryptographic settings. This work has been presented at CCS'18 [24].

Some modern e-voting systems take into account that the platform used for voting may be corrupted, e.g. infected by malware, yet aiming to ensure privacy and integrity of votes even in that case. Bursuc and Kremer, in collaboration with Dragan (Univ of Surrey) propose a new definition of vote privacy, formalized in the cryptographic model as a computational indistinguishability game. The definition captures both known and novel attacks against several voting schemes, and they propose a scheme that is provably secure in this setting. Moreover the proof is formalized and machine-checked in the EasyCrypt theorem prover [45]. This result is currently under submission for publication.

7.2.2. Analysis of e-voting protocols

Participants: Véronique Cortier, Mathieu Turuani.

Belenios is a voting platform designed by our team in collaboration with the Caramba research group at Inria Nancy. Cortier, in collaboration with Warinschi (Univ Bristol), Dragan and Dupressoir (Univ of Surrey), has developed a machine-checked security proof of both privacy and verifiability of Belenios, in the computational model. For this, a novel framework has been developed for proving strong verifiability in EasyCrypt. In the process, several aspects of the pen-and-paper proof of Belenios have been clarified, such as how to deal with revote policies. The framework and the security proofs have been presented at CSF'18 [21].

Turuani and Cortier, in collaboration with Galindo (Univ Birmingham), have analysed the e-voting protocol developed by the Scytl company and planned to be deployed in Switzerland. The formal analysis of both privacy and individual verifiability has been conducted in ProVerif. It required the development of 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 presented at EuroSP'18 [22].

7.2.3. Design of e-voting protocols

Participants: Véronique Cortier, Alicia Filipiak, Joseph Lallemand.

Most existing voting systems either assume trust in the voting device or in the voting server. Filipiak, Lallemand, and Cortier proposed a novel Internet voting scheme, BeleniosVS, that achieves both privacy and verifiability against a dishonest voting server as well as a dishonest voting device. In particular, a voter does not leak her vote to her voting device and she can check that her ballot on the bulletin board does correspond to her intended vote. Additionally, our scheme guarantees receipt-freeness against an external adversary. A formal proof of privacy, receipt-freeness, and verifiability has been established using the tool ProVerif, covering a hundred cases of threat scenarios. Proving verifiability required the identification of a set of sufficient conditions, that can be handled by ProVerif [47]. This contribution is of independent interest. This work is part of the PhD thesis [10] of Alicia Filipiak, defended in March 2018. A conference paper is under submission.

7.3. Privacy

7.3.1. Privacy Protection in Social Networks

Participants: Younes Abid, Bizhan Alipour, Sourya Joyee De, Abdessamad Imine, Michaël Rusinowitch.

To increase awareness about privacy threats, we have designed a tool, SONSAI, for Facebook users to audit their own profiles. SONSAI predicts values of sensitive attributes by machine learning and identifies user public attributes that have guided the learning algorithm towards these sensitive attribute values. The tool is designed to perform reasonably with the limited resources of a personal computer, by collecting and processing only a small relevant part of the network data [31], [32]. We also show how SONSAI is fully interfaced with Facebook along different scenarios. In each case a dataset was built from real profiles collected in the user's neighbourhood network. The whole analysis process is performed online, mostly automatically and with an accuracy of 0.79 when inferring political orientation. More details on the inference of other sensitive attributes are given in [8]. We are now investigating potential privacy attacks based on other data types such as posts, comments and images.

Online social network profiles help users to build new friendships as well as reviving and enhancing existing ones. However, users can become the victims of privacy harms such as identity theft, stalking or discrimination due to the personal data revealed in these profiles. So they have to carefully select the privacy settings for their profile attributes, keeping in mind this trade-off between privacy and social benefit. To aid in this decision process, we have developed a user-friendly model based on Integer Programming [27]. Our model provides a social network user with easy-to-implement suggestions about the privacy settings of his profile attributes such that he can achieve the maximum social benefit while protecting himself from all or at least some major privacy risks. We have tested our approach on user profiles with varying vicinities (i.e. the list of friends) and social benefit requirements [25].

Users' interactions must consider both privacy risks and social benefits, a view supported by the EU General Data Protection Regulation (GDPR). In addition, the GDPR recognizes user consent as a legitimate ground for data processing. In [26], we analyze the present status of user consent in online social networks and we observe that evaluating the privacy risks of user consents to data processing activities can be an effective way to help users in their decision to give or refuse consent.

7.3.2. Compressed and Verifiable Filtering Rules in Software-defined Networking

Participants: Ahmad Abboud, Michaël Rusinowitch.

In a joint project with the Resist research group at Inria Nancy and the Cynapsys/Numeryx companies, we are working on the design, implementation and evaluation of a double-mask technique for building compressed and 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. Fine-Grained Control over Tracking to Support the Ad-Based Web Economy

Participant: Claude Castelluccia.

The intrusiveness of Web tracking and the increasing invasiveness of digital advertising have raised serious concerns regarding user privacy and Web usability, leading a substantial chunk of the populace to adopt ad-blocking technologies in recent years. The problem with these technologies, however, is that they are extremely limited and radical in their approach, and they completely disregard the underlying economic model of the Web, in which users get content free in return for allowing advertisers to show them ads. Nowadays, with around 200 million people regularly using such tools, said economic model is in danger. In this article, we investigate an Internet technology that targets users who are not, in general, against advertising, accept the trade-off that comes with the "free" content, but—for privacy concerns—they wish to exert fine-grained control over tracking. Our working assumption is that some categories of web pages (e.g., related to health or religion) are more privacy-sensitive to users than others (e.g., about education or science). Capitalizing on this, we propose a technology that allows users to specify the categories of web pages that are privacy-sensitive to them and block the trackers present on such web pages only. As tracking is prevented by blocking network connections of third-party domains, we avoid not only tracking but also third-party ads. Since users continue receiving ads on those web pages that belong to non-sensitive categories, our approach may provide a better point of operation within the trade-off between user privacy and the Web economy. To test the appropriateness and feasibility of our solution, we implemented it as a Web-browser plug-in, which is currently available for Google Chrome and Mozilla Firefox. Experimental results from the collected data of 746 users during one year show that only 16.25% of ads are blocked by our tool, which seems to indicate that the economic impact of the ad-blocking exerted by privacy-sensitive users could be significantly reduced.

6.2. 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 this paper, we present 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.

6.3. On the Cost-Effectiveness of Mass Surveillance

Participant: Claude Castelluccia.

In recent times, we have witnessed an increasing concern by governments and intelligence agencies to deploy mass-surveillance systems that help them fight terrorism. Although a government may be perfectly legitimate to do so, it is questionable whether a preventive-surveillance state is rational and cost-effective. In this paper, we conduct a theoretical analysis of the cost of such surveillance systems. Our analysis starts with a fairly well-known result in statistics, namely, the false-positive paradox. We propose a quantitative measure of the total cost of a monitoring program, and study a detection system that is designed to minimize it, subject to a constraint in the percentage of terrorists the agency wishes to capture. Our formulation is first illustrated by means of several simple albeit insightful examples of terrorist and innocent profiles. Then, we conduct an extensive experimental study from real-world socio-demographic data of jihadist terrorism in the U.K. and Spain, and provide insight into the rationality and cost-effectiveness of two countries with two of the biggest defense budgets in the world.

6.4. To Extend or not to Extend: on the Uniqueness of Browser Extensions and Web Logins

Participants: Claude Castelluccia, Gabor Gulyas.

Recent works showed that websites can detect browser extensions that users install and websites they are logged into. This poses significant privacy risks, since extensions and Web logins that reflect user's behavior, can be used to uniquely identify users on the Web. This paper reports on the first large-scale behavioral uniqueness study based on 16,393 users who visited our website. We test and detect the presence of 16,743 Chrome extensions, covering 28% of all free Chrome extensions. We also detect whether the user is connected to 60 different websites. We analyze how unique users are based on their behavior, and find out that 54.86% of users that have installed at least one detectable extension are unique; 19.53% of users are unique among those who have logged into one or more detectable websites; and 89.23% are unique among users with at least one extension and one login. We use an advanced fingerprinting algorithm and show that it is possible to identify a user in less than 625 milliseconds by selecting the most unique combinations of extensions. Because privacy extensions contribute to the uniqueness of users, we study the trade-off between the amount of trackers blocked by such extensions and how unique the users of these extensions are. We have found that privacy extensions should be considered more useful than harmful. The paper concludes with possible counter- measures.

6.5. Privacy-Preserving Release of Spatio-Temporal Density

Participants: Claude Castelluccia, Gergely Acs.

In today's digital society, increasing amounts of contextually rich spatio-temporal information are collected and used, e.g., for knowledge-based decision making, research purposes, optimizing operational phases of city management, planning infrastructure networks, or developing timetables for public transportation with an increasingly autonomous vehicle fleet. At the same time, however, publishing or sharing spatio-temporal data, even in aggregated form, is not always viable owing to the danger of violating individuals' privacy, along with the related legal and ethical repercussions. In this chapter, we review some fundamental approaches for anonymizing and releasing spatio-temporal density, i.e., the number of individuals visiting a given set of locations as a function of time. These approaches follow different privacy models providing different privacy guarantees as well as accuracy of the released anonymized data. We demonstrate some sanitization (anonymization) techniques with provable privacy guarantees by releasing the spatio-temporal density of Paris, in France. We conclude that, in order to achieve meaningful accuracy, the sanitization process has to be carefully customized to the application and public characteristics of the spatio-temporal data.

6.6. Algorithmic Decision Systems in the Health and Justice Sectors: Certification and Explanations for Algorithms in European and French Law

Participant: Daniel Le Metayer.

Algorithmic decision systems are already used in many everyday tools and services on the Internet, and they also play an increasing role in many situations in which people's lives and rights are strongly affected, such as job and loans applications, but also medical diagnosis and therapeutic choices, or legal advice and court decisions. This evolution gives rise to a whole range of questions. In this paper, we argue that certification and explanation are two complementary means of strengthening the European legal framework and enhancing trust in algorithmic decision systems. The former can be seen as the delegation of the task of checking certain criteria to an authority, while the latter allows the stakeholders themselves (for example, developers, users and decision-subjects) to understand the results or the logic of the system. We explore potential legal requirements of accountability in this sense and their effective implementation. These two aspects are tackled from the perspective of the European and French legal frameworks. We focus on two particularly sensitive application domains, namely the medical and legal sectors.

6.7. Capacity: an Abstract Model of Control over Personal Data

Participant: Daniel Le Metayer.

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 propose a formal framework based on capacities to specify the notion of control over personal data and to reason about control properties. We illustrate our framework with social network systems and show 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.

6.8. Biometric Systems Private by Design: Reasoning about privacy properties of biometric system architectures

Participant: Daniel Le Metayer.

In is to show the applicability of the privacy by design approach to biometric systems and the benefit of using formal methods to this end. We build on a general framework for the definition and verification of privacy architectures introduced at STM 2014 and show how it can be adapted 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. Some architectures have already been analysed but on a case by case basis, which makes it difficult to draw comparisons and to provide a rationale for the choice of specific options. In this paper, we describe the application of a general privacy architecture framework to specify different design options for biometric systems and to reason about them in a formal way.

6.9. Privacy Risk Analysis to Enable Informed Privacy Settings

Participant: Daniel Le Metayer.

is a contribution to enhancing individual control over personal data which is promoted, inter alia, by the new EU General Data Protection Regulation. We propose a method to enable better informed choices of privacy preferences or privacy settings. 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.

6.10. Enhancing Transparency and Consent in the IoT

Participants: Daniel Le Metayer, Claude Castelluccia, Mathieu Cunche, Victor Morel.

The development of the IoT raises specific questions in terms of privacy, especially with respect to information to users and consent. We argue that (1) all necessary information about collected data and the collecting devices should be communicated electronically to all data subjects in their range and (2) data subjects should be able to reply also electronically and express their own privacy choices. In this position paper, we take some examples of technologies and initiatives to illustrate our position (including direct and registry-based communications) and discuss them in the light of the GDPR and the WP29 recommendations.

6.11. Toward privacy in IoT mobile devices for activity recognition

Participant: Antoine Boutet.

Recent advances in wireless sensors for personal healthcare allow to recognise human real-time activities with mobile devices. While the analysis of those datastream can have many benefits from a health point of view, it can also lead to privacy threats by exposing highly sensitive information. In this work, we propose a privacy-preserving framework for activity recognition. This framework relies on a machine learning technique to efficiently recognise the user activity pattern, useful for personal healthcare monitoring, while limiting the risk of re-identification of users from biometric patterns that characterizes each individual. To achieve that, we first deeply analysed different features extraction schemes in both temporal and frequency domain. We show that features in temporal domain are useful to discriminate user activity while features in frequency domain lead to distinguish the user identity. On the basis of this observation, we second design a novel protection mechanism that processes the raw signal on the user's smartphone and transfers to the application server only the relevant features unlinked to the identity of users. In addition, a generalisation-based approach is also applied on features in frequency domain before to be transmitted to the server in order to limit the risk of re-identification. We extensively evaluate our framework with a reference dataset: results show an accurate activity recognition (87%) while limiting the re-identification rate (33%). This represents a slightly decrease of utility (9%) against a large privacy improvement (53%) compared to state-of-the-art baselines, while reducing the computational cost on the application server.

6.12. The Long Road to Computational Location Privacy: A Survey

Participant: Antoine Boutet.

The widespread adoption of continuously connected smartphones and tablets developed the usage of mobile applications, among which many use location to provide geolocated services. These services provide new prospects for users: getting directions to work in the morning, leaving a check-in at a restaurant at noon and checking next day's weather in the evening are possible right from any mobile device embedding a GPS chip. In these location-based applications, the user's location is sent to a server, which uses them to provide contextual and personalised answers. However, nothing prevents the latter from gathering, analysing and possibly sharing the collected information, which opens the door to many privacy threats. Indeed, mobility data can reveal sensitive information about users, among which one's home, work place or even religious and political preferences. For this reason, many privacy-preserving mechanisms have been proposed these last years to enhance location privacy while using geolocated services. This work surveys and organises contributions in this area from classical building blocks to the most recent developments of privacy threats and location privacy-preserving mechanisms. We divide the protection mechanisms between online and offline use cases, and organise them into six categories depending on the nature of their algorithm. Moreover, this work surveys the evaluation metrics used to assess protection mechanisms in terms of privacy, utility and performance. Finally, open challenges and new directions to address the problem of computational location privacy are pointed out and discussed.

6.13. CYCLOSA: Decentralizing Private Web Search Through SGX-Based Browser Extensions

Participant: Antoine Boutet.

By regularly querying Web search engines, users (unconsciously) disclose large amounts of their personal data as part of their search queries, among which some might reveal sensitive information (e.g. health issues, sexual, political or religious preferences). Several solutions exist to allow users querying search engines while improving privacy protection. However, these solutions suffer from a number of limitations: some are subject to user re-identification attacks, while others lack scalability or are unable to provide accurate results. This contribution presents CYCLOSA, a secure, scalable and accurate private Web search solution. CYCLOSA improves security by relying on trusted execution environments (TEEs) as provided by Intel SGX. Further, CYCLOSA proposes a novel adaptive privacy protection solution that reduces the risk of user re-identification. CYCLOSA sends fake queries to the search engine and dynamically adapts their count according to the sensitivity of the user query. In addition, CYCLOSA meets scalability as it is fully decentralized, spreading the load for distributing fake queries among other nodes. Finally, CYCLOSA achieves accuracy of Web search as it handles the real query and the fake queries separately, in contrast to other existing solutions that mix fake and real query results.

6.14. ACCIO: How to Make Location Privacy Experimentation Open and Easy

Participant: Antoine Boutet.

The advent of mobile applications collecting and exploiting the location of users opens a number of privacy threats. To mitigate these privacy issues, several protection mechanisms have been proposed this last decade to protect users' location privacy. However, these protection mechanisms are usually implemented and evaluated in monolithic way, with heterogeneous tools and languages. Moreover, they are evaluated using different methodologies, metrics and datasets. This lack of standard makes the task of evaluating and comparing protection mechanisms particularly hard. In this work, we present ACCIO, a unified framework to ease the design and evaluation of protection mechanisms. Thanks to its Domain Specific Language, ACCIO allows researchers and practitioners to define and deploy experiments in an intuitive way, as well as to easily collect and analyse the results. ACCIO already comes with several state-of-the-art protection mechanisms and a toolbox to manipulate mobility data. Finally, ACCIO is open and easily extensible with new evaluation metrics and protection mechanisms. This openness, combined with a description of experiments through a user-friendly DSL, makes ACCIO an appealing tool to reproduce and disseminate research results easier. In this work, we present ACCIO's motivation and architecture, and demonstrate its capabilities through several use cases involving multiples metrics, state-of-the-art protection mechanisms, and two real-life mobility datasets collected in Beijing and in the San Francisco area.

6.15. Collaborative Filtering Under a Sybil Attack: Similarity Metrics do Matter!

Participant: Antoine Boutet.

Recommendation systems help users identify interesting content, but they also open new privacy threats. In this contribution, we deeply analyze the effect of a Sybil attack that tries to infer information on users from a user-based collaborative-filtering recommendation systems. We discuss the impact of different similarity metrics used to identity users with similar tastes in the trade-off between recommendation quality and privacy. Finally, we propose and evaluate a novel similarity metric that combines the best of both worlds: a high recommendation quality with a low prediction accuracy for the attacker. Our results, on a state-of-the-art recommendation framework and on real datasets show that existing similarity metrics exhibit a wide range of behaviors in the presence of Sybil attacks, while our new similarity metric consistently achieves the best trade-off while outperforming state-of-the-art solutions.

6.16. Automatic Privacy and Utility Preservation of Mobility Data: A Nonlinear Model-Based Approach

Participant: Antoine Boutet.

The widespread use of mobile devices and location-based services has generated massive amounts of mobility databases. While processing these data is highly valuable, privacy issues can occur if personal information is revealed. The prior art has investigated ways to protect mobility data by providing a large range of Location Privacy Protection Mechanisms (LPPMs). However, the privacy level of the protected data significantly varies depending on the protection mechanism used, its configuration and on the characteristics of the mobility data. Meanwhile, the protected data still needs to enable some useful processing. To tackle these issues, in this work we present PULP, a framework that finds the suitable protection mechanism and automatically configures it for each user in order to achieve user-defined objectives in terms of both privacy and utility. PULP uses nonlinear models to capture the impact of each LPPM on data privacy and utility levels. Evaluation of our framework is carried out with two protection mechanisms of the literature and four real-world mobility datasets. Results show the efficiency of PULP, its robustness and adaptability. Comparisons between LPPMs' configurator and the state of the art further illustrate that PULP better realizes users' objectives and its computations time is in orders of magnitude faster.

6.17. Privacy Preserving Analytics

Participant: Mathieu Cunche.

As communications-enabled devices are becoming more ubiquitous, it becomes easier to track the movements of individuals through the radio signals broadcasted by their devices. Thus, 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 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 called Pan-Private BLIP. We also extend and generalize previous approaches for estimating distinct count of events and joint events (i.e., intersection and more generally t-out-of-n cardinalities). Finally, we evaluate expirementally our approach and compare it to previous ones on real datasets.

6.18. Detecting smartphone state changes through a Bluetooth based timing attack

Participants: Mathieu Cunche, Guillaume Celosia.

Bluetooth is a popular wireless communication technology that is available on most mobile devices. Although Bluetooth includes security and privacy preserving mechanisms, we show that a Bluetooth harmless inherent request-response mechanism can taint users privacy. More specifically, we introduce a timing attack that can be triggered by a remote attacker in order to infer information about a Bluetooth device state. By observing the L2CAP layer ping mechanism timing variations, it is possible to detect device state changes, for instance when the device goes in or out of the locked state. Our experimental results show that change point detection analysis of the timing allows to detect device state changes with a high accuracy. Finally, we discuss applications and countermeasures.

6.19. Analyzing Ultrasound-based Physical Tracking Systems

Participant: Mathieu Cunche.

A trending application of ultrasound communication is the implementation of ultrasound beacons to track owners of mobile phones in stores and shopping centers. We present the analysis of an Ultrasound-based tracking application. By analyzing several mobile applications along with the network communication and sample of the original audio signal, we were able to reverse engineer the ultrasonic communications and some other elements of the system. Based on those finding we show how arbitrary ultrasonic signal can be generated and how to perform jamming. Finally we analyze a real world deployment and discuss privacy implications.

PROSECCO Project-Team

7. New Results

7.1. Composition Theorems for CryptoVerif and Application to TLS 1.3

Participant: Bruno Blanchet.

We presented composition theorems for security protocols, to compose a key exchange protocol and a symmetric-key protocol that uses the exchanged key. Our results rely on the computational model of cryptography and are stated in the framework of the tool CryptoVerif. They support key exchange protocols that guarantee injective or non-injective authentication. They also allow random oracles shared between the composed protocols. To our knowledge, they are the first composition theorems for key exchange stated for a computational protocol verification tool, and also the first to allow such flexibility.

As a case study, we applied our composition theorems to a proof of TLS 1.3 Draft-18. This work fills a gap in our previous analysis of TLS 1.3 in CryptoVerif [52]. It appears in [31], [39].

7.2. Mechanised Cryptographic Proof of the WireGuard VPN Protocol

Participants: Benjamin Lipp, Bruno Blanchet, Karthikeyan Bhargavan.

WireGuard is a free and open source Virtual Private Network (VPN) that aims to replace IPsec and OpenVPN. It is based on a new cryptographic protocol derived from the Noise Protocol Framework. We provide the first mechanised cryptographic proof of the protocol underlying WireGuard, using the CryptoVerif proof assistant.

We analyse the entire WireGuard protocol as it is, including transport data messages, in an ACCE-style model. We contribute proofs for correctness, message secrecy, forward secrecy, mutual authentication, session uniqueness, and resistance against key compromise impersonation, identity mis-binding, and replay attacks. We also discusse the strength of the identity hiding provided by WireGuard.

Our work also provides novel theoretical contributions that are reusable beyond WireGuard. First, we extend CryptoVerif to account for the absence of public key validation in popular Diffie-Hellman groups like Curve25519, which is used in many modern protocols including WireGuard. To our knowledge, this is the first mechanised cryptographic proof for any protocol employing such a precise model. Second, we prove several indifferentiability lemmas that are useful to simplify the proofs for sequences of key derivations. This work is under submission.

7.3. Meta-F*: Proof automation with SMT, Tactics, and Metaprograms

Participants: Guido Martinez, Danel Ahman, Victor Dumitrescu, Nick Giannarakis [Princeton University], Chris Hawblitzel [Microsoft Research], Catalin Hritcu, Monal Narasimhamurthy [University of Colorado Boulder], Zoe Paraskevopoulou [Princeton University], Clément Pit-Claudel [MIT], Jonathan Protzenko [Microsoft Research], Tahina Ramananandro [Microsoft Research], Aseem Rastogi [Microsoft Research], Nikhil Swamy [Microsoft Research].

We introduced Meta-F* [69], a tactics and metaprogramming framework for the F* program verifier. The main novelty of Meta-F* is allowing to use tactics and metaprogramming to discharge assertions not solvable by SMT, or to just simplify them into well-behaved SMT fragments. Plus, Meta-F* can be used to generate verified code automatically.

Meta-F* is implemented as an F* effect, which, given the powerful effect system of F*, heavily increases code reuse and even enables the lightweight verification of metaprograms. Metaprograms can be either interpreted, or compiled to efficient native code that can be dynamically loaded into the F* type-checker and can interoperate with interpreted code. Evaluation on realistic case studies shows that Meta-F* provides substantial gains in proof development, efficiency, and robustness.

7.4. When Good Components Go Bad: Formally Secure Compilation Despite Dynamic Compromise

Participants: Carmine Abate, Arthur Azevedo de Amorim [CMU], Roberto Blanco, Ana Nora Evans [University of Virginia], Guglielmo Fachini [Nozomi Networks], Catalin Hritcu, Théo Laurent, Benjamin C. Pierce [University of Pennsylvania], Marco Stronati [Nomadic Labs], Andrew Tolmach [Portland State University].

We proposed a new formal criterion [47] for evaluating secure compilation schemes for unsafe languages, expressing end-to-end security guarantees for software components that may become compromised after encountering undefined behavior—for example, by accessing an array out of bounds.

Our criterion is the first to model dynamic compromise in a system of mutually distrustful components with clearly specified privileges. It articulates how each component should be protected from all the others—in particular, from components that have encountered undefined behavior and become compromised. Each component receives secure compilation guarantees—in particular, its internal invariants are protected from compromised components—up to the point when this component itself becomes compromised, after which we assume an attacker can take complete control and use this component's privileges to attack other components. More precisely, a secure compilation chain must ensure that a dynamically compromised component cannot break the safety properties of the system at the target level any more than an arbitrary attacker-controlled component (with the same interface and privileges, but without undefined behaviors) already could at the source level.

To illustrate the model, we construct a secure compilation chain for a small unsafe language with buffers, procedures, and components, targeting a simple abstract machine with built-in compartmentalization. We give a careful proof (mostly machine-checked in Coq) that this compiler satisfies our secure compilation criterion. Finally, we show that the protection guarantees offered by the compartmentalized abstract machine can be achieved at the machine-code level using either software fault isolation or a tag-based reference monitor.

7.5. The Meaning of Memory Safety

Participants: Arthur Azevedo de Amorim [CMU], Catalin Hritcu, Benjamin C. Pierce [University of Pennsylvania].

We give a rigorous characterization of what it means for a programming language to be memory safe [51], capturing the intuition that memory safety supports local reasoning about state. We formalize this principle in two ways. First, we show how a small memory-safe language validates a noninterference property: a program can neither affect nor be affected by unreachable parts of the state. Second, we extend separation logic, a proof system for heap-manipulating programs, with a memory-safe variant of its frame rule. The new rule is stronger because it applies even when parts of the program are buggy or malicious, but also weaker because it demands a stricter form of separation between parts of the program state. We also consider a number of pragmatically motivated variations on memory safety and the reasoning principles they support. As an application of our characterization, we evaluate the security of a previously proposed dynamic monitor for memory safety of heap-allocated data.

7.6. Recalling a Witness: Foundations and Applications of Monotonic State

Participants: Danel Ahman, Cédric Fournet [Microsoft Research], Catalin Hritcu, Kenji Maillard, Aseem Rastogi [Microsoft Research], Nikhil Swamy [Microsoft Research].

We provide a way to ease the verification of programs whose state evolves monotonically [48]. The main idea is that a property witnessed in a prior state can be soundly recalled in the current state, provided (1) state evolves according to a given preorder, and (2) the property is preserved by this preorder. In many scenarios, such monotonic reasoning yields concise modular proofs, saving the need for explicit program invariants. We distill our approach into the monotonic-state monad, a general yet compact interface for Hoare-style reasoning about monotonic state in a dependently typed language. We prove the soundness of the monotonic-state monad and use it as a unified foundation for reasoning about monotonic state in the F* verification system. Based on this foundation, we build libraries for various mutable data structures like monotonic references and apply these libraries at scale to the verification of several distributed applications.

7.7. A Monadic Framework for Relational Verification: Applied to Information Security, Program Equivalence, and Optimizations

Participants: Niklas Grimm [Vienna University of Technology], Kenji Maillard, Cédric Fournet [Microsoft Research], Catalin Hritcu, Matteo Maffei [Vienna University of Technology], Jonathan Protzenko [Microsoft Research], Tahina Ramananandro [Microsoft Research], Aseem Rastogi [Microsoft Research], Nikhil Swamy [Microsoft Research], Santiago Zanella-Béguelin [Microsoft Research].

Relational properties describe multiple runs of one or more programs. They characterize many useful notions of security, program refinement, and equivalence for programs with diverse computational effects, and they have received much attention in the recent literature. Rather than developing separate tools for special classes of effects and relational properties, we advocate using a general purpose proof assistant as a unifying framework for the relational verification of effectful programs. The essence of our approach is to model effectful computations using monads and to prove relational properties on their monadic representations, making the most of existing support for reasoning about pure programs [67].

We apply this method in F* and evaluate it by encoding a variety of relational program analyses, including information flow control, program equivalence and refinement at higher order, correctness of program optimizations and game-based cryptographic security. By relying on SMT-based automation, unary weakest preconditions, user-defined effects, and monadic reification, we show that, compared to unary properties, verifying relational properties requires little additional effort from the F* programmer.

7.8. A Formal Treatment of Accountable Proxying over TLS

Participants: Karthikeyan Bhargavan, Ioana Boureanu [University of Surrey], Antoine Delignat-Lavaud [Microsoft Research], Pierre-Alain Fouque [University of Rennes], Cristina Onete [University of Limoges].

Much of Internet traffic nowadays passes through active proxies, whose role is to inspect, filter, cache, or transform data exchanged between two endpoints. To perform their tasks, such proxies modify channel-securing protocols, like TLS, resulting in serious vulnerabilities. Such problems are exacerbated by the fact that middleboxes are often invisible to one or both endpoints, leading to a lack of accountability. A recent protocol, called mcTLS, pioneered accountability for proxies, which are authorized by the endpoints and given limited read/write permissions to application traffic.

Unfortunately, we show that mcTLS is insecure: the protocol modifies the TLS protocol, exposing it to a new class of middlebox-confusion attacks. Such attacks went unnoticed mainly because mcTLS lacked a formal analysis and security proofs. Hence, our second contribution is to formalize the goal of accountable proxying over secure channels. Third, we propose a provably-secure alternative to soon-to-be-standardized mcTLS: a generic and modular protocol-design that carefully composes generic secure channel-establishment protocols, which we prove secure. Finally, we present a proof-of-concept implementation of our design, instantiated with unmodified TLS 1.3 draft 23, and evaluate its overheads [29].

7.9. hacspec: towards verifiable crypto standards

Participants: Karthikeyan Bhargavan, Franziskus Kiefer [Mozilla], Pierre-Yves Strub [Ecole Polytechnique].

We designed and published hacspec, a formal specification language for cryptographic primitives. Specifications (specs) written in hacspec are succinct, easy to read and implement, and lend themselves to formal verification using a variety of existing tools. The syntax of hacspec is similar to the pseudocode used in cryptographic standards but is equipped with a static type system and syntax checking tools that can find errors. Specs written in hacspec are executable and can hence be tested against test vectors taken from standards and specified in a common format. Finally, hacspec is designed to be compilable to other formal specification languages like F*, EasyCrypt, Coq, and cryptol, so that it can be used as the basis for formal proofs of functional correctness and cryptographic security using various verification frameworks.

We published a paper presenting the syntax, design, and tool architecture of hacspec. We demonstrated the use of the language to specify popular cryptographic algorithms, and developed preliminary compilers from hacspec to F* and to EasyCrypt. Our eventual goal is to invite authors of cryptographic standards to write their pseudocode in hacspec and to help the formal verification community develop the language and tools that are needed to promote high-assurance cryptographic sofware backed by mathematical proofs. All our code is released publicly on GitHub.

7.10. Largest-scale user study of secure messaging and API usage

Participants: Francesca Musiani [CNRS], Ksenia Ermoshina [CNRS], Harry Halpin, Iness Ben Guirat [INSAT].

As part of the NEXTLEAP EC project, we engaged in the largest ever user study of secure messaging applications, focusing on typical users as well as “high-risk” users in the Middle East and Ukraine, as well as developers.[41]. This work has been shared with standardization efforts such as the IETF Message Layer Security (MLS) effort in which Inria is participating, as well as W3C standardization of the W3C Web Authentication API. This work helped influence the formal verification of the privacy properties of hardware-based cryptographic authentication, which is a feature needed by many at risk users whose accounts are often the focus of hacks. This work has also led a fundamental inquiry into the social governance of standards and the role of formal verification in the future of standards.[42] As this work is highly interdisciplinary, it has featured collaboration with sociologists at CNRS and interns from INSAT in Tunisia, as well as a lecture series hosted at Centre Pompidou under the direction of Bernard Stiegler and Harry Halpin.

TAMIS Project-Team

7. New Results

7.1. Results for Axis 1: Vulnerability analysis

7.1.1. Statistical Model Checking of Incomplete Stochastic Systems

Participants: Tania Richmond, Louis-Marie Traonouez, Axel Legay.

We proposed a statistical analysis of stochastic systems with incomplete information. These incomplete systems are modelled using discrete time Markov chains with unknowns (qDTMC), and the required behaviour was formalized using qBLTL logic. By doing both quantitative and qualitative analysis of such systems using statistical model checking, we also proposed refinement on the qDTMCs. These refined qDTMCs depict a decrease in the probability of unknown behaviour in the system. The algorithms for both qualitative and quantitative analysis of qDTMC were implemented in the tool Plasma Lab. We demonstrated the working of these algorithms on a case study of a network with unknown information. We plan to extend this work to analyse the behaviour of other stochastic models like Markov decision processes and abstract Markov chains, with incomplete information.

This work has been accepted and presented to a conference this year [10].

[10] We study incomplete stochastic systems that are missing some parts of their design, or are lacking information about some components. It is interesting to get early analysis results of the requirements of these systems, in order to adequately refine their design. In previous works, models for incomplete systems are analysed using model checking techniques for three-valued temporal logics. In this paper, we propose statistical model checking algorithms for these logics. We illustrate our approach on a case-study of a network system that is refined after the analysis of early designs.

7.1.2. A Language for Analyzing Security of IOT Systems

Participants: Delphine Beaulaton, Najah Ben Said, Ioana Cristescu, Axel Legay, Jean Quilbeuf.

We propose a model-based security language of Internet of Things (IoT) systems that enables users to create models of their IoT systems and to make analysis of the likelihoods of cyber-attacks to occur and succeed. The modeling language describes the interactions between different entities, that can either be humans or “Things” (i.e, hardware, sensors, software tools, ..). A malicious entity is present in the system, called the Attacker, and it carries out attacks against the system. The other IoT entities can inadvertently help the Attacker, by leaking their sensitive data. Equipped with the acquired knowledge the Attacker can then communicate with the IoT entities undetected. For instance, an attacker can launch a phishing attack via email, only if it knows the email address of the target.

Another feature of our modeling language is that security failures are modeled as a sequence of simpler steps, in the spirit of *attack trees*. As their name suggests, attacks are modeled as trees, where the leaves represent elementary steps needed for the attack, and the root represents a successful attack. The internal nodes are of two types, indicating whether all the sub-goals (an AND node) or one of the sub-goals (an OR node) must be achieved in order to accomplish the main goal. The attack tree provided with the IoT system acts as a monitor: It observes the interactions the Attacker has with the system and detects when an attack is successful.

An IoT system is analyzed using statistical model checking (SMC). The first method we use is Monte Carlo, which consists of sampling the executions of an IoT system and computing the probability of a successful attack based on the number of executions for which the attack was successful. However, the evaluation may be difficult if a successful attack is *rare*. We therefore also use a second SMC method, developed for *rare events*, called *importance splitting*.

To implement this we rely on *BIP*, a heterogeneous component-based model for which an execution engine is developed and maintained. The IoT model is translated into a BIP model and the attack tree into a BIP monitor. The two form a BIP system. The execution engine of BIP produce executions which are the input of Plasma Lab, the model checker developped in TAMIS. We have extended Plasma Lab with a plugin that interacts with the BIP execution engine.

The tools are available at http://iot-modeling.gforge.inria.fr/. This work has been published in two conference papers [20], [23]. A third paper was submitted in November [29], and is currently under review.

[20] In this paper we propose our security-based modeling language for IoT systems. The modeling language has two important features: (i) vulnerabilities are explicitly represented and (ii) interactions are allowed or denied based on the information stored on the IoT devices. An IoT system is transformed in BIP, a component-based modeling language, in which can execute the system and perform security analysis. To illustrate the features of our language, we model a use-case based on a Smart Hospital and inspired by industrial scenarios.

[23] In this paper we revisit the security-based modeling language for IoT systems. We focus here on the BIP models obtained from the original IoT systems. The BIP execution and analysis framework provides several methods to analyse a BIP model, and we discuss how these methods can be lifted on the original IoT systems. We also model a new use-case based on Amazon Smart Home.

[29] Attack trees are graphical representations of the different scenarios that can lead to a security failure. In this paper we extend our security-based framework for modeling IoT systems in two ways: (i) attack trees are defined alongside the model to detect and prevent security risks in the system and (ii) the language supports probabilistic models. A successful attack can be a *rare event* in the execution of a well designed system. When rare, such attacks are hard to detect with usual model checking techniques. Hence, we use *importance splitting* as a statistical model checking technique for rare events.

7.1.3. Verification of IKEv2 protocol

Participants: Tristan Ninet, Olivier Zendra, Louis-Marie Traonouez, Axel Legay.

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. IKEv2 guarantees security properties like mutual-authentication and secrecy of exchanged key. To obtain an IKEv2 implementation as secure as possible, we use model checking to verify the properties on the protocol specification, and software formal verification tools to detect implementation flaws like buffer overflows or memory leaks.

In previous analyses, IKEv2 has been shown to possess two authentication vulnerabilities that were considered not exploitable. We analyze the protocol specification using the Spin model checker, and prove that in fact the first vulnerability does not exist. In addition, we show that the second vulnerability is exploitable by designing and implementing a novel slow Denial-of-Service attack, which we name the Deviation Attack.

We propose an expression of the time at which Denial-of-Service happens, and validate it through experiment on the strongSwan implementation of IKEv2. As a counter-measure, we propose a modification of IKEv2, and use model checking to prove that the modified version is secure.

For ethical reasons we informed our country's national security agency (ANSSI) about the existence of the Deviation Attack. The security agency gave us some technical feedback as well as its approval for publishing the attack.

We then tackle formal verification directly applied to an IKEv2 source code. We already tried to analyze strongSwan using the Angr tool. However we found that the Angr was not mature yet for a program like strongSwan. We thus try other software formal verification tools and apply them to smaller and simpler source code than strongSwan: we analyze OpenSSL asn1parse using the CBMC tool and light-weight IP using the Infer tool. We find that CBMC does not scale to a large source code and that Infer does not verify the properties we want.

We plan to explore more in-depth a formal technique and work towards the goal of verifying generic properties (absence of implementation flaws) on softwares like strongSwan.

7.1.4. Combining Software-based and Hardware-based Fault Injection Approaches

Participants: Nisrine Jafri, Annelie Heuser, Jean-Louis Lanet, Axel Legay, Thomas Given-Wilson.

Software-based and hardware-based approaches have both been used to detect fault injection vulnerabilities. Software-based approaches can provide broad and rapid coverage as it was shown in the previous publications [36], [37], [38] , but may not correlate with genuine hardware vulnerabilities. Hardware-based approaches are indisputable in their results, but rely upon expensive expert knowledge and manual testing.

This work bridges software-based and hardware-based fault injection vulnerability detection by contrasting results of both approaches. To our knowledge no research where done trying to bridge the software-based and hardware-based approach to detect fault injection vulnerabilities the way it is done in this work.

Using both the software-based and hardware-based approaches showed that:

- Software-based approaches detect genuine fault injection vulnerabilities.
- Software-based approaches yield false-positive results.
- Software-based approaches did *not* yield false-negative results.
- Not all software-based vulnerabilities can be reproduced in hardware.
- Hardware-based EMP approaches do *not* have a simple fault model.
- There is a coincidence between software-based and hardware-based approaches.
- Combining software-based and hardware-based approaches yields a vastly more efficient method to detect genuine fault injection vulnerabilities.

This work implemented both the SimFI tool and the ArmL tool.

7.1.5. Side-channel analysis on post-quantum cryptography

Participants: Annelie Heuser, Tania Richmond.

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.

We started by a survey on cryptanalysis against code-based cryptography [13], that includes algebraic and side-channel attacks. Code-based cryptography reveals sensitive data mainly in the syndrome decoding. We investigate the syndrome computation from a side-channel point of view. There are different methods that can be used depending on the underlying code. We explore vulnerabilities of each one in order to propose a guideline for designers and developers. This work was presented at CryptArchi 2018 and Journées Codes et Cryptographie 2018.

[13] Nowadays public-key cryptography is based on number theory problems, such as computing the discrete logarithm on an elliptic curve or factoring big integers. Even though these problems are considered difficult to solve with the help of a classic computer, they can be solved in polynomial time on a quantum computer. Which is why the research community proposed alternative solutions that are quantum resistant. The process of finding adequate post-quantum cryptographic schemes has moved to the next level, right after NIST's announcement for post-quantum standardization.

One of the oldest quantum resistant proposition goes back to McEliece in 1978, who proposed a public-key cryptosystem based on coding theory. It benefits of really efficient algorithms as well as strong mathematical backgrounds. Nonetheless, its security has been challenged many times and several variants were cryptanalyzed. However, some versions are still unbroken.

In this paper, we propose to give a short background on coding theory in order to present some of the main flawless in the protocols. We analyze the existing side-channel attacks and give some recommendations on how to securely implement the most suitable variants. We also detail some structural attacks and potential drawback for new variants.

7.1.6. New Advances on Side-channel Distinguishers

Participants: Christophe Genevey Metat, Annelie Heuser, Tania Richmond.

[17] *On the Performance of Deep Learning for Side-channel Analysis* We answer the question whether convolutional neural networks are more suitable for SCA scenarios than some other machine learning techniques, and if yes, in what situations. Our results point that convolutional neural networks indeed outperforms machine learning in several scenarios when considering accuracy. Still, often there is no compelling reason to use such a complex technique. In fact, if comparing techniques without extra steps like preprocessing, we see an obvious advantage for convolutional neural networks only when the level of noise is small, and the number of measurements and features is high. The other tested settings show that simpler machine learning techniques, for a significantly lower computational cost, perform similar or even better. The experiments with the guessing entropy metric indicate that simpler methods like Random forest or XGBoost perform better than convolutional neural networks for the datasets we investigated. Finally, we conduct a small experiment that opens the question whether convolutional neural networks are actually the best choice in side-channel analysis context since there seems to be no advantage in preserving the topology of measurements.

[8] *The Curse of Class Imbalance and Conflicting Metrics with Machine Learning for Side-channel Evaluations* We concentrate on machine learning techniques used for profiled side-channel analysis in the presence of imbalanced data. Such scenarios are realistic and often occurring, for instance in the Hamming weight or Hamming distance leakage models. In order to deal with the imbalanced data, we use various balancing techniques and we show that most of them help in mounting successful attacks when the data is highly imbalanced. Especially, the results with the SMOTE technique are encouraging, since we observe some scenarios where it reduces the number of necessary measurements more than 8 times. Next, we provide extensive results on comparison of machine learning and side-channel metrics, where we show that machine learning metrics (and especially accuracy as the most often used one) can be extremely deceptive. This finding opens a need to revisit the previous works and their results in order to properly assess the performance of machine learning in side-channel analysis.

[35] *When Theory Meets Practice: A Framework for Robust Profiled Side-channel Analysis* Profiled side-channel attacks are the most powerful attacks and they consist of two steps. The adversary first builds a leakage model, using a device similar to the target one, then it exploits this leakage model to extract the secret information from the victim's device. These attacks can be seen as a classification problem, where the adversary needs to decide to what class (corresponding to the secret key) the traces collected from the victim's devices belong to. For a number of years, the research community studied profiled attacks and proposed numerous improvements. Despite a large number of empirical works, a framework with strong theoretical foundations to address profiled side-channel attacks is still missing.

In this paper, we propose a framework capable of modeling and evaluating all profiled analysis attacks. This framework is based on the expectation estimation problem that has strong theoretical foundations. Next, we quantify the effects of perturbations injected at different points in our framework through robustness analysis where the perturbations represent sources of uncertainty associated with measurements, non-optimal classifiers, and methods. Finally, we experimentally validate our framework using publicly available traces, different classifiers, and performance metrics.

[33] *Make Some Noise: Unleashing the Power of Convolutional Neural Networks for Profiled Side-channel Analysis* Profiled side-channel attacks based on deep learning, and more precisely Convolutional Neural Networks, is a paradigm showing significant potential. The results, although scarce for now, suggest that such techniques are even able to break cryptographic implementations protected with countermeasures. In this paper, we start by proposing a new Convolutional Neural Network instance that is able to reach high performance for a number of considered datasets. Additionally, for a dataset protected with the random delay countermeasure, our neural network is able to break the implementation by using only 2 traces in the attack phase. We compare our neural network with the one designed for a particular dataset with masking countermeasure and we show how both are good designs but also how neither can be considered as a superior to the other one. Next, we address how the addition of artificial noise to the input signal can be actually beneficial to the performance of the neural network. Such noise addition is equivalent to the regularization term in the objective function. By using this technique, we are able to improve the number of measurement needed to reveal the secret key by orders of magnitude in certain scenarios for both neural networks. To strengthen our experimental results, we experiment with a number of datasets which differ in the levels of noise (and type of countermeasure) where we show the viability of our approaches.

[9] *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.

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: Olivier Decourbe, Annelie Heuser, Jean-Louis Lanet, Olivier Zendra, Cassius Puodzius, Stefano Sebastio, Lamine Nourredine, Jean Quilbeuf, Eduard Baranov, Thomas Given-Wilson, Fabrizio Biondi, Axel Legay, Alexander Zhdanov.

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: Olivier Decourbe, Lamine Nourredine, Annelie Heuser, Nisrine Jafri, Jean-Louis Lanet, Jean Quilbeuf, Axel Legay, Fabrizio Biondi.

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.

- **1. Packing Detection and Classification**

  Packing is a widespread tool to prevent static malware detection and analysis. Detecting and classifying the packer used by a given malware sample is fundamental to being able to unpack and study the malware, whether manually or automatically. Existing works on packing detection and classification has focused on effectiveness, but does not consider the efficiency required to be part of a practical malware-analysis workflow. This work studies how to train packing detection and classification algorithms based on machine learning to be both highly effective and efficient. Initially, we create ground truths by labeling more than 280,000 samples with three different techniques. Then we perform feature selection considering the contribution and computation cost of features. Then we iterate over more than 1,500 combinations of features, scenarios, and algorithms to determine which algorithms are the most effective and efficient, finding that a reduction of 1-2% effectiveness can increase efficiency by 17-44 times. Then, we test how the best algorithms perform against malware collected after the training data to assess them against new packing techniques and versions, finding a large impact of the ground truth used on algorithm robustness. Finally, we perform an economic analysis and find simple algorithms with small feature sets to be more economical than complex algorithms with large feature sets based on uptime/training time ratio.

- **2. Packing clustering** A limit of supervised learning is to not be able to recognize classes that were not present in the ground truth. In the work's case above, this means that packer families for which a classifier has not been trained will not be recognized. In this work, we use unsupervised learning techniques, more particularly clustering, in order to provide information about packed malware with previously unknown packing techniques. Here, we build our own dataset of packed binaries, since in the previous work, it has been shown that the construction of the ground truth was fundamental in determining the effectiveness of the packing classification process. Choosing the right clustering algorithm with the right distance metric, dealing with different scales of features units, while being effective, efficient and robust are also majors parts of the current work.

This work is still in progress ...

- **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 and clustering

Participants: Annelie Heuser, Nisrine Jafri, Jean-Louis Lanet, Cassius Puodzius, Stefano Sebastio, Olivier Decourbe, Eduard Baranov, Jean Quilbeuf, Thomas Given-Wilson, Axel Legay, Fabrizio Biondi.

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.

Moreover, we explored the impact on the malware classification of several heuristics adoptable in the SCDGs building process and graph exploration. In particular, our purpose was to:

- identify quality characteristics and evaluation metrics of binary signatures based on SCDGs (and consequently the key properties of the execution traces), that characterize signatures able to provide high-precision malware classification
- optimize the performance of the SMT solver by designing a meta-heuristic able to select the best heuristic to tackle a specific sub-class of problem, study the impact of the configuration of the SMT solver and symbolic execution framework, and understand their interdependencies with the aim of efficiently extracting SCDGs in accordance with the identified quality metrics.

By adopting a Design of Experiments approach constituted by a full factorial experiment design and an Analysis of Variance (ANOVA) we have been able to pinpoint that, considering the graph metrics and their impact on the F-score, the litmus test for the quality of an SCDG-based classifier is represented by the presence of connected components. This could be explained considering how the graph mining algorithm (gSpan) works and the adopted similarity metric based on the number of common edges between the extracted signatures and the SCDG of the sample to classify. The results of the factorial experiments show that in our context tuning the symbolic execution is a very complex problem and that the sparsity of effect principle (stating that the system is dominated by the effect of the main factors and low-order-factor interactions) does not hold. The evaluation proved that the SMT solver is the most influential positive factor also showing an ability in reducing the impact of heuristics that may need to be enabled due to resource constraints (e.g., the max number of active paths). Results suggest that the most important factors are the disjoint union (as trace combination heuristic), and the our SMT optimization (through meta-heuristics) whereas other heuristics (such as min trace size and step timeout) have less impact on the quality of the constructed SCDGs.

Preliminary experiments show the promising results of our approach by considering the F-score in the classification of the malware families. Further investigation are needed in particular by using a larger dataset. For this purpose we established an academic collaboration with VirusTotal for helping us to build a ground truth for the family name.

One fundamental issue for supervised learning is the trustworthiness of the settled ground truth. In the scenario of malware classification, it is common to have great disagreement in the labeling of the very same malware sample (e.g. family attributed by different anti-malware vendors). Therefore, unsupervised learning on malware datasets by clustering based on the similarities of their SCDGs allows to overcome this problem.

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. Papers

This section gathers papers that are results common to all sections above pertaining to Axis 2.

- Efficient Extraction of Malware Signatures Through System Calls and Symbolic Execution: An Experience Report [28]

The ramping up use of network connected devices is providing hackers more incentives and opportunities to design and spread new security threats. Usually, malware analysts employ a mix of automated tools and human expertise to study the behavior of suspicious binaries and design suitable countermeasures. The analysis techniques adopted by automated tools include symbolic execution.Symbolic execution envisages the exploration of all the possible execution paths of the binary without neither concretizing the values of the variables nor dynamically executing the code (i.e., the binary is analyzed statically). Instead, all the values are represented symbolically. Progressing in the code exploration, constraints on symbolic variables are built and system calls tracked. A satisfiability-modulo-theory (SMT) checker is in charge of verifying the satisfiability of the collected symbolic constraints and thus the validity of an execution path. Unfortunately, while widely considered promising, this approach suffers from high resource consumption. Therefore, optimizing the constraint solver and tuning the features controlling symbolic execution is of fundamental importance to effectively adopting the technique. In this paper, we identify the metrics characterizing the quality of binary signatures expressed as system call dependency graphs extracted from a malware database. Then, we pinpoint some optimizations allowing to extract better binary signatures and thus to outperform the vanilla version of symbolic analysis tools in terms of malware classification and exploitation of the available resources.

7.3. Other research results

7.3.1. ContAv: a Tool to Assess Availability of Container-Based Systems

Participant: Stefano Sebastio.

This work was the result of a collaboration with former members of XRCI (Xerox Research Centre India): Rahul Ghosh, Avantika Gupta and Tridib Mukherjee.

[18] (C) The momentum gained by the microservice-oriented architecture is fostering the diffusion of operating system containers. Existing studies mainly focus on the performance of containerized services to demonstrate their low resource footprints. However, availability analysis of densely deployed container-based solutions is less visited due to difficulties in collecting failure artifacts. This is especially true when the containers are combined with virtual machines to achieve a higher security level. Inspired by Google's Kubernetes architecture, in this paper, we propose ContAv, an open-source distributed statistical model checker to assess availability of systems built on containers and virtual machines. The availability analysis is based on novel state-space and non-state-space models designed by us and that are automatically built and customized by the tool. By means of a graphical interface, ContAv allows domain experts to easily parameterize the system, to compare different configurations and to perform sensitivity analysis. Moreover, through a simple Java API, system architects can design and characterize the system behavior with a failure response and migration service.

7.3.2. (Coordination of the) TeamPlay Project, and Expression of Security Properties

Participants: Olivier Zendra, Yoann Marquer, Céline Minh, Annelie Heuser, Tania Richmond.

This work is done in the context of the TeamPlay EU project.

As mobile applications, the Internet of Things, and cyber-physical systems become more prevalent, so there is an increasing focus on energy efficiency of multicore computing applications. At the same time, traditional performance issues remain equally important. Increasingly, software designs need to find the best performance within some energy budget, often while also respecting real-time or other constraints, which may include security, data locality or system criticality, and while simultaneously optimising the usage of the available hardware resources.

While parallel multicore/manycore hardware can, in principle, ameliorate energy problems, and heterogeneous systems can help to find a good balance between execution time and energy usage, at present there are no effective analyses beyond user-guided simulations that can reliably predict energy usage for parallel systems, whether alone or in combination with timing information and security properties. In order to create energy-, time- and security- (ETS) efficient parallel software, programmers need to be actively engaged in decisions about energy usage, execution time and security properties rather than passively informed about their effects. This extends to design-time as well as to implementation-time and run-time.

In order to address this fundamental challenge, TeamPlay takes a radically new approach: by exploiting new and emerging ideas that allow non-functional properties to be deeply embedded within their programs, programmers can be empowered to directly treat energy ETS properties as first-class citizens in their parallel software. The concrete objectives of the TeamPlay project are:

1. To develop new mechanisms, along with their theoretical and practical underpinnings, that support direct language-level reasoning about energy usage, timing behaviour, security, etc.
2. To develop system-level coordination mechanisms that facilitate optimised resource usage for multicore hardware, combining system-level resource utilisation control during software development with efficient spatial and temporal scheduling at run-time.
3. To determine the fundamental inter-relationships between time, energy, security, etc. optimisations, to establish which optimisation approaches are most effective for which criteria, and to consequently develop multiobjective optimising compilers that can balance energy consumption against timing and other constraints.
4. To develop energy models for heterogeneous multicore architectures that are sufficiently accurate to enable high-level reasoning and optimisation during system development and at run-time.
5. To develop static and dynamic analyses that are capable of determining accurate time, energy usage and security information for code fragments in a way that can inform high-level programs, so achieving energy, time and security transparency at the source code level.
6. To integrate these models, analyses and tools into an analysis-based toolbox that is capable of reflecting accurate static and dynamic information on execution time and energy consumption to the programmer and that is capable of optimising time, energy, security and other required metrics at the whole system level.
7. To identify industrially-relevant metrics and requirements and to evaluate the effectiveness and potential of our research using these metrics and requirements.
8. To promote the adoption of advanced energy-, time- and security-aware software engineering techniques and tools among the relevant stake-holders.

Inria will exploit the results of the TeamPlay project in two main domains. First, they will strengthen and extend the research Inria has been carrying on low power and energy for embedded systems, especially for memory and wireless sensors networks. Second, they will complement in a very fitting way the research carried at Inria about security at a higher level (model checking, information theory).

The capability to express the energy and security properties at the developper level will be integrate in Inria own prototype tools, hence widening their applicability and the ease of experimentation. The use of energy properties wrt. evening of energy consumption to prevent information leakage, thus making side-channels attacks more difficult, is also a very promising path.

In addition, the methodological results pertaining to the development of embedded systems with a focus on low power and energy should also contribute to research lead at Inria in the domain of software engineering and advanced software engineering tools. Furthermore, security research lead at Inria will benefit from the security work undertaken by Inria and SIC in TeamPlay.

Overall, the project, with a strong industrial presence, will allow Inria to focus on matching concrete industrial requirements aiming at actual products, hence in providing more robust and validated results. In addition, the extra experience of working with industrial partners including SMEs will surely impact positively on Inria research methodology, making Inria research more attractive and influential, especially wrt. industry.

Finally, the results, both in terms of methodology and techniques, will also be integrated in the teaching Inria contributes to at Master level, in the areas of Embedded Systems and of Security.

The TeamPlay consortium agreement has been created by Inria, discussed with the various partners, and has been signed by all partners on 28 Feb. 2018. Inria has also distributed the partners initial share of the grant at the beginning of the project.

As WP7 (project management) leader and project coordinator, Inria was in charge of arranging general project meetings, including monthly meetings (tele-conferences), bi-annual physical meetings, boards meetings. During the first period, three exceptional physical meetings have been conducted, in addition to monthly project meetings: the kick-off meeting in Rennes from the 30th to the 31st of January 2018, the physical progress meeting has been conducted in Odense from the 26th to the 27th of June 2018, and the review in Brussels prepared the 19th of September 2018 and set the 17th of October 2018.

We have selected and set up utility tools for TeamPlay: shared notepads, mailing lists, shared calendars and collaborative repositories. We have ensured the timely production of the due deliverables. We set up the Project Advisory Board (PAB) with the aim of gathering external experts from both academia and industry, covering a wide range of domains addressed by TeamPlay. Finally, we ensured good working relationships (which can implicate conflict resolution when needed), monitored the overall progress of the project, and reported to the European Commission on technical matters and deliverables.

We also organized a tooling meeting in Hamburg in October the 30th, to discuss the relation between the tools from different partners, e.g. Idris from the University of St Andrews, the WCC compiler developed in the Hamburg University of Technology, or the coordination tool developed in the University of Amsterdam.

Measuring security, unlike measuring other more common non-functional properties like time or energy, is still very much in its infancy. For example, time is often measured in seconds (or divisions thereof), but security has no widely agreed, well-defined measurement. It is thus one goal of this project, especially for SIC and Inria, to design (necessarily novel) security measurements, and have them implemented as much as possible throughout the set of development tools.

Measuring security by only one value however seems impossible or may be meaningless. More precisely, if security could be defined overall by only one measurement, the latter would be a compound (i.e. an aggregation) of several more specialized measurement. Indeed, security encompasses many aspects of interest:

1. By allowing communications between different systems, security properties should be guaranteed in order to prevent low-level users from determining anything about high-level users activity, or in the case of public communication channels in a hostile environment, to evaluate vulnerability to intruders performing attacks on communications.
    1. *Confidentiality* (sometimes called *secrecy*) properties like non-interference (and many) variants can be described by using an information-flow policy (e.g. high- and low-level users) and studying traces of user inputs.
    2. *Vulnerability* captures how a system is sensible to attacks on communications (e.g. stealing or faking information on a public channel).
2. A *side-channel* is a way of transmitting informations (purposely or not) to another system out of the standard (intended) communication channels. *Side-channel attacks* rely on the relationship between information leaked through a side-channel and the secret data to obtain confidential (non-public) information.
    1. *Entropy* captures the uncertainty of the attacker about the secret key. The attacker must be able to extract information about the secret key through side-channel measurements, which is captured by the *attacker's remaining uncertainty* value, which can be computed by using heuristic techniques. The attacker must also be able to effectively recover the key from the extracted information, which is expressed by the *min-entropy leakage*, and refined by the *g-leakage* of a gain function.
    2. The power consumption of a cryptographic device can be analyzed to extract the secret key. This is done by using several techniques: visual examination of graphs of the current (*Simple Power Analysis*), by exploiting biases in varying power consumption (*Differential Power Analysis*), or by using the correlation coefficient between the power samples and hypotheses (*Correlation Power Analysis*).

   3. Usual security properties guarantee only the input-output behavior of a program, and not its execution time. Closing *leakage through timing* can be done by disallowing while-loops and if-commands to depend on high security data, or by padding the branches so that the external observer cannot determine which branch was taken.
   4. Finally, the correlation between the patterns of the victim's execution and the attacker's observations is formalized as a metric called the *Side-channel Vulnerability Factor*, which is refined by the *Cache Side-channel Vulnerability* for cache attacks.

3. A cryptographic scheme should be secure even if the attacker knows all details about the system, with the exception of the secret keys. In particular, the system should be secure when the attacker knows the encryption and decryption algorithms.
   1. In modern cryptography, the security level (or security strength) is given by the *work factor*, which is related to its key-length and the number of operations necessary to break a cryptographic scheme (try all possible combinations of the key). An algorithm is said to have a "security level of n bits" if the best known attack requires 2^n steps. This is a quite natural definition because symmetric algorithms with a security level of n have a key of length n bits.
   2. The relationship between cryptographic strength and security is not as straightforward in the asymmetric case. Moreover, for symmetric algorithms, a key-length of 128 bits provides an estimated long term security (i.e. several decades in the absence of quantum computer) regarding brute-force attacks. To reach an estimated long term security even with quantum computers, a key-length of 256 bits is mandatory.

Inria is implementing side-channel countermeasures (hiding) into the WCET-aware C Compiler (WCC) developed by the Hamburg University of Technology (TUHH). A research visit to TUHH was arranged with the aim at learning how to work on WCC (TUHH and WCC infrastructure, WCC developers best practices, etc.). Inria will use compiler-based techniques to prevent timing leakages and power leakages.

For instance, in a conditional branching `if` b `then` $P_1(x)$ `else` $P_2(x)$, measuring the execution time or the power profile may allow to know whether the branch P_1 or P_2 have been chosen to manipulate the value x, thus to obtain the secret value b. To prevent timing leakage, P_1 and/or P_2 can be padded (i.e. dummy instructions are added) in order to obtain the worst-case execution time in both branches.

But this does not prevent information leakage from power profile. A stronger technique, from a security point of view, could be to add a dummy variable y and duplicate the code such that $y = x;$`if` b `then` $P_1(x); P_2(y)$ `else` $P_1(Y); P_2(x)$ always performs the operations of P_1 then the operations of P_2. But the execution time is now the sum and not the worst-case of both branches, thus trading execution time to increase security.

Finally, the initialization $y = x$ can be detected, and the previous solution is still vulnerable to fault injections. Some algorithms like the Montgomery Ladder are more protected against these attacks because both variables x and y are entangled during the execution. We hope to generalize this property to a wider set of algorithms, or to automatically detect the properties required from the original code in order to transform it into a “Montgomerised" version with higher security level.